Bounded Predicates in Description Logics with Counting

Sanja Lukumbuzya and **Mantas Šimkus**

Institute of Logic and Computation, TU Wien, Austria
lukumbuzya@kr.tuwien.ac.at, simkus@dbai.ac.at

Abstract

Description Logics (DLs) support so-called *anonymous objects*, which significantly contribute to the expressiveness of these KR languages, but also cause substantial computational challenges. This paper investigates reasoning about upper bounds on predicate sizes for ontologies written in the expressive DL $\mathcal{ALCHOIQ}$ extended with *closed predicates*. We describe a procedure based on integer programming that allows us to decide the existence of upper bounds on the cardinality of some predicate in the models of a given ontology in a data-independent way. Our results yield a promising supporting tool for constructing higher quality ontologies, and provide a new way to push the decidability frontiers. To wit, we define a new safety condition for Datalog-based queries over DL ontologies, while retaining decidability of query entailment.

1 Introduction

Knowledge Representation and Reasoning (KR&R) is playing an increasingly important role in intelligent *data management*, especially for complex and knowledge-intensive problem domains. Specifically, KR&R offers *Description Logics (DLs)* as languages suitable for describing complex domains in terms of *ontologies*, and automated reasoning tools allow us to infer useful insights from data repositories leveraging such ontologies. DLs are expressive enough to capture and reason about UML and ER diagrams, which are popular modeling languages in the development of data-centric applications [Berardi *et al.*, 2005]. The adequate expressiveness and good computational properties of DLs have also led to the emergence of the so-called *Ontology-based Data Access (OBDA)* paradigm for information integration; see [Xiao *et al.*, 2018; Schneider and Šimkus, 2020] for an overview of this area.

A distinguishing feature of DLs, aimed at dealing with information incompleteness, is the ability to describe and reason about *anonymous objects*, that is, elements in the domain of interest that are not represented by known individuals but whose existence is logically implied by the background knowledge. Consider a simple TBox with two axioms Employee $\equiv$ {Robin} $\sqcup$ {Skyler} and Employee $\sqsubseteq$ $\neg$Task $\sqcap$ $\geq$2hasTask.Task $\sqcap$ $\leq$5hasTask.Task, which tells that Robin and Skyler are precisely the employees of a company, that employees are not tasks, and that each employee must be assigned from 2 to 5 tasks to work on. If we inspect the models of this TBox, then the *named* objects corresponding to Robin and Skyler will be associated with 2 to 5 *anonymous* objects corresponding to tasks.

When the number of anonymous objects cannot be bounded (e.g., when it is forced to be infinite by a recursive TBox), one often faces high computational complexity of standard reasoning tasks, and undecidability of more sophisticated problems, like answering Datalog-based recursive queries (see, e.g., [Levy and Rousset, 1998; Rosati, 2007]) or reasoning about data-manipulating actions (see, e.g., [Bagheri Hariri *et al.*, 2013]). However, in many scenarios, the number of anonymous objects in relevant predicates can be inferred to be bounded due to numeric constraints present in the TBox. E.g., if we add to the above TBox the statement that every task must be associated to an employee (Task $\sqsubseteq$ $\exists$hasTask$^-$.Employee), then the extension of Task in any model is bounded by 10. We suggest [Bednarczyk *et al.*, 2020] for a longer discussion of numeric constraints in DLs.

The question we investigate in this paper is the existence of upper bounds on predicate sizes for ontologies written in the very expressive DL $\mathcal{ALCHOIQ}$ with *closed predicates* [Franconi *et al.*, 2011; Lutz *et al.*, 2013]. In the context of relational databases, closed predicates are related to the notion of *master data*, corresponding to the static part of a database [Fan and Geerts, 2010]. E.g., in a database with a (frequently updated) table of sales transactions, the supporting tables of employees, products and clients will likely be relatively static and thus can be considered master data. Using a TBox, closed and non-closed (open) predicates can be related by means of inclusion axioms, analogously to the way that sales transactions can be related to employees, products and customers via database dependency constraints. Our goal is to understand when and how we can infer bounds on the sizes of open predicates from the extensions of closed predicates, by taking into account the (numeric and other) constraints specified in a TBox. Our results not only yield a promising tool for supporting the construction of high-quality ontologies, but also open a new way to push the decidability frontiers for challenging data management tasks. Our contributions can be summarized as follows.

∘ We introduce the notion of *bounded predicates* for DLs with closed predicates. Intuitively, a predicate is bounded in

a knowledge base, if there exists an integer constant that provides an upper bound on the size of the predicate's extension in all models of the knowledge base. Since we are interested in aiding the design of ontologies and of the so-called *ontology-mediated queries (OMQs)*, in the paper we consider a stronger variant of predicate boundedness that requires the predicate to be bounded independently of the actual extensional data stored in a knowledge base. More precisely, we require the existence of a function f such that $f(n)$ is an upper bound for all ABoxes of size at most n.

 ○ We formalize two decision problems – checking predicate boundedness in the weak and in the strong sense, and we characterize their computational complexity. For $\mathcal{ALCHOIQ}$, our method is based on integer programming and it involves an intermediate step that reformulates the problem in terms of another problem called *finite-infinite satisfiability*. The task in this problem is to check whether a TBox has a model in which some specific predicates have *finite extensions*, while some other given predicates have *infinite extensions*. This is closely related to *mixed satisfiability* studied in [Gogacz *et al.*, 2020a] and is interesting in its own right.

 ○ We show that our results yield worst-case optimal complexity bounds in all cases. Specifically, checking boundedness in the weak and the strong sense is CONEXPTIME-complete for $\mathcal{ALCHOIQ}$ TBoxes. Moreover, in case predicate boundedness is inferred, the concrete bound can be readily computed; it is double exponential in the size of the input. This is worst-case optimal: one can craft a TBox that forces a double exponential number of elements in a bounded predicate. However, in case of strong boundedness, the function f as mentioned above is a polynomial function, i.e. if we fix the TBox, the number of elements in bounded predicates will grow polynomially in the size of the input ABox.

 ○ Finally, we show how our results can be used to define a new decidability-ensuring safety condition for OMQs based on $\mathcal{ALCHOIQ}$ and Datalog, providing worst-case optimal results on the combined and data complexity. This is interesting because there are very few positive results on query answering in this expressive DL. We also infer that the data complexity of satisfiability of $\mathcal{ALCHOIQ}$ KBs with closed predicate is CONP-complete.

An extended version of this paper can be found at https://dbai.tuwien.ac.at/staff/simkus/papers/ijcai21-bnd.pdf

2 Preliminaries

We now introduce $\mathcal{ALCHOIQ}$, the main DL investigated in the paper. We assume countably infinite and mutually disjoint sets $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ of *concept names*, *role names*, and *constants*, respectively. *Roles* are expressions of the form p and p^-, where $p \in \mathsf{N_R}$. We let $\mathsf{N_R^+}$ be the set of all roles. With a slight abuse of notation, we write r^- to denote p^- if $r = p$, and p if $r = p^-$, for $p \in \mathsf{N_R}$. The role r^- is the *inverse* of r. Given a set R of roles, R^- denotes the set $\{r^- : r \in R\}$. $\mathcal{ALCHOIQ}$ *concepts* are defined according to the following syntax: $C := \top \mid \bot \mid A \mid \{a\} \mid \neg C \mid C \sqcup C \mid C \sqcap C \mid \forall r.C \mid \leq nr.C \mid \geq nr.C \mid = nr.C$, where $A \in \mathsf{N_C}$, r is a role, $a \in \mathsf{N_I}$ and $n \geq 0$. Concepts of the form $\{a\}$, where $a \in \mathsf{N_I}$

are *nominals*. An *ABox* is a finite set of *assertions* of the form $A(a)$ or $p(a, b)$, where $a, b \in \mathsf{N_I}$, $A \in \mathsf{N_C}$, and $p \in \mathsf{N_R}$. Given $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$, $\mathcal{A}|_\Sigma$ denotes the restriction of $\mathcal{A}$ to the concept and role names in Σ. If Σ consists of a single concept or role name p, we simply write $\mathcal{A}|_p$ to denote $\mathcal{A}|_\Sigma$. An expression $C \sqsubseteq D$, where C and D are concepts, is a *concept inclusion*, and an expression $r \sqsubseteq s$, where r and s are roles, is a *role inclusion*. A *(TBox) axiom* is a concept inclusion or a role inclusion, and a *TBox* is a finite set of axioms. A *knowledge base (with closed predicates)* (KB) is a triple $(\mathcal{T}, \Sigma, \mathcal{A})$, where $\mathcal{T}$ is a TBox, $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$ is a set of closed predicates and $\mathcal{A}$ is an ABox. For a TBox, an ABox, or a KB $\mathcal{X}$, we denote by $\mathsf{N_C}(\mathcal{X})$ and $\mathsf{N_R}(\mathcal{X})$ the set of concept and role names occurring in $\mathcal{X}$, and by $\mathsf{N_R^+}(\mathcal{X})$ the set of roles occurring in $\mathcal{X}$ and their inverses. As $\mathcal{ALCHOIQ}$ is simply a fragment of first-order logic with a special syntax, its semantics is given in term of first-order interpretations. An interpretation is a pair $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set called the *domain* and $\cdot^\mathcal{I}$ is the *interpretation function* that assigns to each $a \in \mathsf{N_I}$ a domain element $a^\mathcal{I} \in \Delta^\mathcal{I}$, to each $A \in \mathsf{N_C}$ a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, and to each $r \in \mathsf{N_R}$ a set $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The extension of the interpretation function to the remaining concepts and roles is defined in the standard way [Baader *et al.*, 2003]. A concept or a role inclusion $Q_1 \sqsubseteq Q_2$ is satisfied by an interpretation $\mathcal{I}$ if $Q_1^\mathcal{I} \subseteq Q_2^\mathcal{I}$. We note that we make the *Standard Name Assumption (SNA)*, which is common when dealing with closed predicates and which forces us to interpret every constant as itself, i.e., $a^\mathcal{I} = a$, for all $\mathcal{I}$ and all constants a. We say that $\mathcal{I}$ satisfies a TBox $\mathcal{T}$ (resp. ABox $\mathcal{A}$) if it satisfies all axioms in $\mathcal{T}$ (resp. assertions in $\mathcal{A}$). $\mathcal{I}$ satisfies a KB $(\mathcal{T}, \Sigma, \mathcal{A})$ if $\mathcal{I}$ satisfies $\mathcal{T}$ and $\mathcal{A}$, and $\{p(\vec{a}) : p \in \Sigma, \vec{a} \in p^\mathcal{I}\} = \mathcal{A}|_\Sigma$.

3 Boundedness

We begin with an example that illustrates boundedness.

Example 1. *Consider the following $\mathcal{ALCHOIQ}$ TBox $\mathcal{T} = \{\mathsf{Empl} \sqsubseteq \leq 5\ \mathsf{assgndTo.Proj}, \mathsf{Proj} \sqsubseteq \geq 1\mathsf{assgndTo^-.Empl}\}$ stating that each employee in some company can be assigned to at most 5 projects, and that all projects must be assigned to least one employee. Let $\mathcal{A}$ be an arbitrary ABox over* Empl *that lists all n employees of the company, i.e., $\mathcal{A} = \{\mathsf{Empl}(a_1), \dots, \mathsf{Empl}(a_n)\}$. Further, let $\mathcal{I}$ be an arbitrary model of $(\mathcal{T}, \{\mathsf{Empl}\}, \mathcal{A})$. As* Empl *is viewed as a closed predicate, $|\mathsf{Empl}^\mathcal{I}| = |\mathcal{A}| = n$. Taking into account the axioms from $\mathcal{T}$, it is easy to see that the number of projects in $\mathcal{I}$ can be at most $5n$, i.e., $|\mathsf{Proj}^\mathcal{I}| \leq 5n$. Therefore, the size of the extensions of* Proj *is in a way bounded w.r.t. $\mathcal{T}$ and* Empl.

Definition 1. *Let $\mathcal{T}$ be a DL TBox, Σ be a set of predicates occurring in $\mathcal{T}$, and $\mathcal{A}$ be an ABox over the signature of $\mathcal{T}$. A predicate p is* bounded *w.r.t $(\mathcal{T}, \Sigma, \mathcal{A})$ if there exists a bound $b \in \mathbb{N}$ s.t. in every model $\mathcal{I}$ of $(\mathcal{T}, \Sigma, \mathcal{A})$, $|p^\mathcal{I}| \leq b$ holds. We say that p is* bounded *w.r.t $\mathcal{T}$ and Σ if p is bounded w.r.t. $(\mathcal{T}, \Sigma, \mathcal{A})$, for every ABox $\mathcal{A}$ over the signature of $\mathcal{T}$.*

The problem of deciding whether all predicates in a given set are bounded w.r.t. a given TBox formulated in the DL $\mathcal{L}$ and a set of closed predicates is called *strong boundedness*, or BOUNDEDNESS($\mathcal{L}$), and it is defined as follows.

BOUNDEDNESS($\mathcal{L}$)

Input: A triple $(\mathcal{T}, \Sigma, \Sigma_B)$, where $\mathcal{T}$ is a TBox in DL $\mathcal{L}$ and Σ, Σ_B are sets of predicates occurring in $\mathcal{T}$.

Question: Is each $C \in \Sigma_B$ bounded w.r.t. $\mathcal{T}$ and Σ?

We also define the second decision problem called *weak boundedness*, or W-BOUNDEDNESS($\mathcal{L}$), that given a TBox $\mathcal{T}$ formulated in the DL $\mathcal{L}$, a set of predicates Σ, and an ABox $\mathcal{A}$ over the signature of $\mathcal{T}$, decides whether all predicates in some given set are bounded w.r.t. $(\mathcal{T}, \Sigma, \mathcal{A})$.

W-BOUNDEDNESS($\mathcal{L}$)

Input: A pair $(\mathcal{K}, \Sigma_B)$, where $\mathcal{K} = (\mathcal{T}, \Sigma, \mathcal{A})$, $\mathcal{T}$ is a TBox in DL $\mathcal{L}$, Σ, Σ_B are sets of predicates occurring in $\mathcal{T}$, and $\mathcal{A}$ is an ABox over the signature of $\mathcal{T}$.

Question: Is each $C \in \Sigma_B$ bounded w.r.t. $\mathcal{K}$?

We now focus specifically on strong boundedness for description logics that are fragments of the first-order logic (FO). Note that, if not specified otherwise, under *boundedness* we understand strong boundedness. As a technical tool we introduce *finite-infinite satisfiability* (*fi-satisfiability*), or FI-SAT($\mathcal{L}$), which is the problem of deciding for a given TBox $\mathcal{T}$ in the DL $\mathcal{L}$, and two sets of predicates Σ_F and Σ_I, whether $\mathcal{T}$ has a model in which all the predicates in Σ_F have finite extensions and all the predicates in Σ_I have infinite extensions.

FI-SAT($\mathcal{L}$)

Input: A triple $(\mathcal{T}, \Sigma_F, \Sigma_I)$, where $\mathcal{T}$ is a TBox in DL $\mathcal{L}$, $\Sigma_F \cup \Sigma_I$ is a set of predicates occurring in $\mathcal{T}$.

Question: Is there a model $\mathcal{I}$ of $\mathcal{T}$ s.t. $p^{\mathcal{I}}$ is finite for all $p \in \Sigma_F$ and is infinite for all $p \in \Sigma_I$?

We next show that we can reduce the boundedness problem to deciding fi-satisfiability. From the Definition 1, it is easy to observe that a concept or a role name p is not bounded w.r.t. $\mathcal{T}$ and Σ if and only if there exists an ABox $\mathcal{A}$ over the signature of $\mathcal{T}$ such that for every natural number n there exists a model $\mathcal{I}$ of $(\mathcal{T}, \Sigma, \mathcal{A})$ with $|p^{\mathcal{I}}| > n$. The following proposition, formulated for FO, shows that in this case $(\mathcal{T}, \Sigma, \mathcal{A})$ also has a model in which the extension of p is infinite.

Proposition 1. *Let $\mathcal{T}$ be an arbitrary FO theory and let p be a unary or a binary predicate. If for every natural number n there exists a model $\mathcal{I}$ of $\mathcal{T}$ in which $|p^{\mathcal{I}}| > n$, then there also exists a model $\mathcal{J}$ of $\mathcal{T}$ in which $p^{\mathcal{J}}$ is infinite.*

Proof. Let p be a binary predicate. For $k \geq 1$, let $\varphi_k := \exists x_1, x'_1, \ldots, x_k, x'_k \bigwedge_{1 \leq i < j \leq k} (x_i \neq x_j \vee x'_i \neq x'_j) \wedge p(x_i, x'_i)$, which is satisfied by interpretations in which there are k or more tuples in the extension of p. Consider the FO theory $\mathcal{T}' = \mathcal{T} \cup \{\varphi_n : n \geq 1\}$. As for every n, there is a model $\mathcal{I}$ of $\mathcal{T}$ in which $|p^{\mathcal{I}}| > n$, every finite subset of $\mathcal{T}'$ is satisfiable. By compactness of FO, $\mathcal{T}'$ is satisfiable as well. However, by construction, $\mathcal{T}'$ only admits models in which the extension of p is infinite. Hence, there exists a model $\mathcal{J}$ of $\mathcal{T}'$ such that $p^{\mathcal{J}}$

is infinite. As $\mathcal{T} \subseteq \mathcal{T}'$, $\mathcal{J}$ is also a model of $\mathcal{T}$ and so there is a model of $\mathcal{T}$ in which the extension of p is infinite. The case where p is unary is shown analogously. $\square$

Proposition 2. *Given a TBox $\mathcal{T}$ and a set of predicates Σ occurring in $\mathcal{T}$, a predicate p is not bounded w.r.t. $\mathcal{T}$ and Σ if and only if there exists a model $\mathcal{I}$ of $\mathcal{T}$ in which all predicates from Σ have finite extensions and $p^{\mathcal{I}}$ is infinite.*

Proof. Assume that p is not bounded w.r.t. $\mathcal{T}$ and Σ. Then there exists an ABox $\mathcal{A}$ s.t. for every $b \in \mathbb{N}$, there exists a model $\mathcal{J}$ of $(\mathcal{T}, \Sigma, \mathcal{A})$ in which $|p^{\mathcal{J}}| > b$. Note that, since we are considering DLs that are fragments of FO, we can obtain from $(\mathcal{T}, \Sigma, \mathcal{A})$ an FO theory $\mathcal{T}'$ whose models coincide with those of $(\mathcal{T}, \Sigma, \mathcal{A})$. We let $\mathcal{T}' = \text{fo}(\mathcal{T}) \cup \mathcal{A} \cup \{\varphi_p : p \in \Sigma\}$, where $\text{fo}(\mathcal{T})$ is a theory obtained by translating $\mathcal{T}$ into FO, and φ_p, for $p \in \Sigma$, encodes the semantics of the closed predicates as follows. For each concept name $A \in \Sigma$, let $\Delta_A = \{c : A(c) \in \mathcal{A}\}$. Let $\varphi_A := \forall x A(x) \rightarrow \bigvee_{c \in \Delta_A} x = c$. For a role name $r \in \Sigma$, φ_r is defined in a similar manner. It is easy to see that $\mathcal{T}'$ has exactly the same models as $(\mathcal{T}, \Sigma, \mathcal{A})$. Due to Proposition 1, there is a model $\mathcal{I}$ of $\mathcal{T}'$ (and thus also of $(\mathcal{T}, \Sigma, \mathcal{A})$) in which $p^{\mathcal{I}}$ is infinite. As ABoxes are finite and the extensions of the predicates in Σ are fully specified by $\mathcal{A}$, we have that the predicates in Σ have finite extensions in $\mathcal{I}$.

Let $\mathcal{J}$ be a model of $\mathcal{T}$ s.t. the predicates in Σ have finite extensions and p has an infinite extension in $\mathcal{J}$, and let $\mathcal{A} = \{q(\vec{c}) : \vec{c} \in q^{\mathcal{I}}, q \in \Sigma\}$. As $\mathcal{J}$ is a model of $(\mathcal{T}, \Sigma, \mathcal{A})$ and $|p^{\mathcal{J}}| > k$, for all $k \in \mathbb{N}$, p is not bounded w.r.t. $\mathcal{T}$ and Σ. $\square$

Corollary 1. *Let $\mathcal{T}$ be a TBox in DL $\mathcal{L}$, and Σ, Σ_B be sets of predicates occurring in $\mathcal{T}$. $(\mathcal{T}, \Sigma, \Sigma_B)$ is a yes-instance of* BOUNDEDNESS($\mathcal{L}$) *if and only if $(\mathcal{T}, \Sigma, \{p\})$ is a no-instance of* FI-SAT($\mathcal{L}$)*, for all $p \in \Sigma_B$.*

4 Boundedness via Integer Programming

This section investigates boundedness in $\mathcal{ALCHOIQ}$. We first show that we can reduce fi-satisfiability to solving a system of linear inequalities with side conditions. We are thus able to show decidability of boundedness in $\mathcal{ALCHOIQ}$ and provide tight complexity bounds. The second part of the section shows that there is a function depending on the TBox and $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$, that for every $n \geq 0$, computes an upper bound on the size of extensions of bounded predicates in the models of $(\mathcal{T}, \Sigma, \mathcal{A})$, for all ABoxes $\mathcal{A}$ of size n.

4.1 Decidability and Complexity

Let $(\mathcal{T}, \Sigma_F, \Sigma_I)$ be an instance of FI-SAT($\mathcal{ALCHOIQ}$). For ease of presentation, we assume that Σ_F and Σ_I contain only concept names. This is not a limitation, as role names can be eliminated from both sets in polynomial time while preserving fi-satisfiability. This is done by introducing a fresh concept name A_r, for each role name $r \in \Sigma_F \cup \Sigma_I$, and adding axioms that ensure A_r collects the domain elements in the domain and range of r. Finally, r is replaced by A_r in Σ_F and Σ_I.

The concepts in $\mathsf{N_C^+} = \mathsf{N_C} \cup \{\top, \bot\} \cup \{\{c\} : c \in \mathsf{N_I}\}$ are called *basic concepts* and the set $\mathsf{N_C^+}(\mathcal{T})$ consists of the basic concepts occurring in a TBox $\mathcal{T}$. We next fix some notation that remains the same for the rest of the paper. For a given

TBox $\mathcal{T}$, we let $n_{\mathcal{T}} = |\mathsf{N_C}(\mathcal{T}) \cup \mathsf{N_R^+}(\mathcal{T})|$, $m_{\mathcal{T}}$ be the number of constants occurring in $\mathcal{T}$ plus 1, and let $c_{\mathcal{T}}$ be the maximum integer occurring in $\mathcal{T}$. We assume that TBoxes are in *normal form* in which all axioms are of the following types:

$$B_1 \sqcap \cdots \sqcap B_{k-1} \sqsubseteq B_k \sqcup \cdots \sqcup B_m,$$
$$B_1 \sqsubseteq = n\, p.B_2, \qquad B_1 \sqsubseteq \forall p.B_2, \qquad r \sqsubseteq s,$$

where $\{B_1, \ldots, B_m\} \subseteq \mathsf{N_C^+}$, $n \geq 0$, $k > 1$, $m \geq k$, $p \in \mathsf{N_R}$, and $\{r, s\} \subseteq \mathsf{N_R^+}$. We call the axioms of type $B_1 \sqsubseteq = n\, r.B_2$ *counting axioms*. We also assume that every TBox $\mathcal{T}$ is closed under role inclusions, i.e., (i) $p \sqsubseteq p \in \mathcal{T}$, for each $p \in \mathsf{N_R}(\mathcal{T})$, (ii) if $r \sqsubseteq s \in \mathcal{T}$ then $r^- \sqsubseteq s^- \in \mathcal{T}$, and (iii) if $r_1 \sqsubseteq r_2 \in \mathcal{T}$ and $r_2 \sqsubseteq r_3 \in \mathcal{T}$ then $r_1 \sqsubseteq r_3 \in \mathcal{T}$. Every $\mathcal{ALCHOIQ}$ TBox can be transformed in polynomial time into a TBox in normal form that is closed under role inclusions.

We are now ready to describe the procedure for deciding fi-satisfiability in $\mathcal{ALCHOIQ}$. Note that fi-satisfiability is a generalization of mixed-satisfiability introduced in [Gogacz *et al.*, 2020a], which is the problem of deciding, for a given TBox $\mathcal{T}$ and a set of predicates Σ, whether there exists a model of $\mathcal{T}$ in which all predicates from Σ have finite extensions. Inspired by the techniques in [Calvanese, 1996; Lutz *et al.*, 2005; Pratt-Hartmann, 2005],[Gogacz *et al.*, 2020a] presents a decision procedure for mixed-satisfiability in $\mathcal{ALCHOIF}$, a DL closely related to $\mathcal{ALCHOIQ}$ that only supports functionality. This procedure relies on encoding mixed-satisfiability as a system of linear integer inequalities with side conditions, and a similar encoding was used in [Gogacz *et al.*, 2020b] for deciding satisfiability of $\mathcal{ALCHOIF}$ KBs with closed predicates. We next adapt the procedure in [Gogacz *et al.*, 2020a] to support our generalized setting.

We use *tiles* to describe different types of domain elements in models of $\mathcal{T}$. Intuitively, a tile specifies basic concepts that an element participates in and the relevant part of its neighborhood. A domain element d in a model of $\mathcal{T}$ is said to be an instance of a tile τ if it fits the description given by τ. We then define a system of integer linear inequalities with side conditions whose solutions assign a multiplicity to each tile in a way that ensures that we can build a desired model of $\mathcal{T}$ by instantiating tiles according to their multiplicities. In order to properly support counting, we employ a technique from [Pratt-Hartmann, 2005] that colors the models of $\mathcal{T}$ using $\lceil log(m_{\mathcal{T}} c_{\mathcal{T}}^2 + 1) \rceil$ fresh concept names. For each of these fresh concept names A, we add an axiom $A \sqsubseteq A \in \mathcal{T}$. We next formally define the notion of tiles.

Definition 2. *Given an $\mathcal{ALCHOIQ}$ TBox $\mathcal{T}$, a type T for $\mathcal{T}$ is any set with the following properties: (i) $T \subseteq \mathsf{N_C^+}(\mathcal{T})$, (ii) $\top \in T$ and $\bot \notin T$, and (iii) $|T \cap \{\{a\} : a \in \mathsf{N_I}(\mathcal{T})\}| \leq 1$.*

Definition 3. *Given an $\mathcal{ALCHOIQ}$ TBox $\mathcal{T}$, a tile for $\mathcal{T}$ is a tuple (T, ρ), where T is a type for $\mathcal{T}$ and ρ is a set of triples (R, T', k) s.t. $R \subseteq \mathsf{N_R^+}(\mathcal{T})$, T' is a type for $\mathcal{T}$, $k > 0$ and the following conditions hold:*

1. *If $(R, T', k) \in \rho$ then $(R, T, k) \in \rho$, for all $0 < k' < k$*

2. *For every $(R, T', k) \in \rho$, there exists some $A \sqsubseteq = nr.B \in \mathcal{T}$ such that $A \in T$, $B \in T'$ and $r \in R$*

3. *If $B_1 \sqcap \cdots \sqcap B_{k-1} \sqsubseteq B_k \sqcup \cdots \sqcup B_m \in \mathcal{T}$ and $\{B_1, \ldots, B_{k-1}\} \subseteq T$, then $\{B_k, \ldots, B_m\} \cap T \neq \emptyset$*

4. *If $A \sqsubseteq = n\, r.B \in \mathcal{T}$ and $A \in T$, then $|\{(R, T', k) \in \rho : r \in R \text{ and } B \in T'\}| = n$.*

5. *For all $(R, T', k) \in \rho$, the following hold:*

   (a) *If $A \sqsubseteq \forall r.B \in \mathcal{T}$, $A \in T$ and $r \in R$, then $B \in T'$*

   (b) *If $A \sqsubseteq \forall r.B \in \mathcal{T}$, $A \in T'$ and $r^- \in R$, then $B \in T$*

   (c) *If $r \sqsubseteq s \in \mathcal{T}$ and $r \in R$, then $s \in R$*

   (d) *If (T, R, T') is invertible, then $T \neq T'$, and there is no other $(R', T', l') \in \rho$ s.t. (T, R', T') is invertible,*

where, for types T, T' and $R \subseteq \mathsf{N_R^+}(\mathcal{T})$, (T, R, T') is invertible if there exists $\{A \sqsubseteq = n\, s.B, C \sqsubseteq = m\, p.D\} \subseteq \mathcal{T}$, s.t. $\{A, D\} \subseteq T$, $\{B, C\} \subseteq T'$, and $\{s, p^-\} \subseteq R$.

Let Tiles$(\mathcal{T})$ denote the set of all tiles for $\mathcal{T}$. A tile (T, ρ) describes a domain element d in some model of $\mathcal{T}$ that participates in the basic concepts in T. Moreover, for each $(R, T', l) \in \rho$ there is a domain element d' whose basic concepts are given by T' s.t. (d, d') is in the extension of every role $r \in R$. We say that d' is an R-successor of d.

Example 2. *Consider the following TBox $\mathcal{T} = \{B \sqcap C \sqsubseteq A, A \sqsubseteq = 1r.B, A \sqsubseteq = 1r.\{b\}, A \sqsubseteq \forall r.C, s \sqsubseteq r\}$. Let*

$$\tau = (\{\top, A\}, \{(\{r, s\}, \{\top, A, B, C, \{b\}\}, 1)\}).$$

Then, τ is a tile for $\mathcal{T}$ and it describes a certain kind of domain elements that may appear in models of $\mathcal{T}$. In a model $\mathcal{I}$ of $\mathcal{T}$, a domain element $d \in \Delta^{\mathcal{I}}$ that is an instance of τ has the following properties:

1. *d is of the type $\{\top, A\}$, i.e., $d \in A^{\mathcal{I}}$ and $d \notin D^{\mathcal{I}}$, for all $D \in \mathsf{N_C^+}(\mathcal{T}) \setminus \{A\}$,*

2. *$(d, d') \in r^{\mathcal{I}} \cap s^{\mathcal{I}}$, where $d' \in \Delta^{\mathcal{I}}$ is of the type $\{\top, A, B, C, \{b\}\}$, i.e., $d' = b$, $d \in A^{\mathcal{I}} \cap B^{\mathcal{I}} \cap C^{\mathcal{I}}$, and $d \notin D^{\mathcal{I}}$, for all $D \in \mathsf{N_C^+}(\mathcal{T}) \setminus \{A, B, C\}$.*

We briefly explain the intuitions behind the conditions in Def. 3. Recall that, for a tile $\tau = (T, \rho)$, each $(R, T', k) \in \rho$ represents a distinct R-successors of the domain element d that is an instance of τ. By definition, ρ is a set, which means no duplicates are allowed. However, a situation could arise in which we need two encode two distinct R-successors of d that have the same type and are connected to the d via the same roles. To overcome this issue, we consider triples (R, T', k), where k is an integer that tells us that (R, T', k) encodes the kth R-successor of d of type T'. This is reflected in the first condition. It is also worth mentioning that tiles only encode the relevant part of the neighborhood of a domain element, i.e., those neighbors that serve as witnesses for counting axioms in TBox $\mathcal{T}$ of the knowledge base. This is reflected in condition 2. The intuition behind the remaining conditions is rather straight forward. Condition 3 ensures the satisfaction of axioms of the type $B_1 \sqcap \cdots \sqcap B_{k-1} \sqsubseteq B_k \sqcup \cdots \sqcup B_m \in \mathcal{T}$, condition 4 guarantees that d has n witnesses for every axiom $A \sqsubseteq = nr.B \in \mathcal{T}$, conditions 5a-5c ensure that d and its neighbors respect axioms of the type $A \sqsubseteq \forall r.B$ and $r \sqsubseteq s$, and condition 5d is related to the coloring of models and intuitively, it ensures that a domain element is not used as a witness for counting axiom more times than it is allowed.

We now introduce *enriched systems of linear integer inequalities* that are tuples $(V, \mathcal{E}, V_F, V_I, I)$, where V is a set of variables, $\mathcal{E}$ is a set of inequalities of the form $a_1 \cdot x_1 + \cdots + a_n \cdot x_n + c \leq b_1 \cdot y_1 + \cdots + b_m \cdot y_m$, s.t. $a_1, \ldots, a_n, b_1, \ldots, b_m$ are positive integers, c is a possibly negative integer, and $x_1, \ldots, x_n, y_1, \ldots, y_m \in V$, $V_F \subseteq V$, $V_I \subseteq 2^V$, and I is the set of implications of the form $y_1 + \cdots + y_m > a \Rightarrow x_1 + \cdots + x_n > b$, where a, b are non-negative integers. Recall that we aim to characterize fi-satisfiability via enriched systems s.t. the solutions to the systems specify how many times we need to instantiate each tile in order to build a desired model. As some tiles might need to be instantiated infinitely many times, we extend the set of natural numbers with a new value $\aleph_0$ greater than any natural number and we let: $\aleph_0 \cdot \aleph_0 = \aleph_0 + \aleph_0 = \aleph_0 + 0 = 0 + \aleph_0 = \aleph_0 + n = n + \aleph_0 = \aleph_0 \cdot n = n \cdot \aleph_0 = \aleph_0$, for all $n \in \mathbb{N} \setminus \{0\}$, and $0 \cdot \aleph_0 = \aleph_0 \cdot 0 = 0$. We let $\mathbb{N}^* = \mathbb{N} \cup \{\aleph_0\}$. Furthermore, we say that infinite sets have cardinality $\aleph_0$. A *solution* of $(V, \mathcal{E}, V_F, V_I, I)$ is a function $S : V \to \mathbb{N}^*$ s.t. all inequalities and implications are satisfied, for every $x \in V_F$, $S(x) \neq \aleph_0$, and for every $X \in V_I$, there exists $x \in X$ s.t. $S(x) = \aleph_0$.

Proposition 3. *Let $\mathcal{T}$ be an $\mathcal{ALCHOIQ}$ TBox, and $\Sigma_F, \Sigma_I \subseteq \mathsf{N}_\mathsf{C}(\mathcal{T})$. We can obtain an enriched system $\mathcal{S}_{(\mathcal{T}, \Sigma_F, \Sigma_I)}$ that is exponential in the size of $\mathcal{T}$ s.t. $V = Tiles(\mathcal{T})$, and $\mathcal{S}_{(\mathcal{T}, \Sigma_F, \Sigma_I)}$ has a solution S if and only if there exists a model $\mathcal{I}$ of $\mathcal{T}$ in which $A^\mathcal{I}$ is finite for all $A \in \Sigma_F$ and infinite for all $A \in \Sigma_I$. Moreover, for all concept names $A \in \mathsf{N}_\mathsf{C}^+(\mathcal{T})$, $|A^\mathcal{I}| = \sum_{\substack{(T, \rho) \in Tiles(\mathcal{T}), \\ A \in T}} S((T, \rho))$ holds.*

Theorem 1. FI-SAT($\mathcal{ALCHOIQ}$) *is* NEXPTIME-*complete.* BOUNDEDNESS($\mathcal{ALCHOIQ}$) *is* CONEXPTIME-*complete.*

4.2 Size of Bounded Extensions

Let $\mathcal{T}$ be an $\mathcal{ALCHOIQ}$ TBox and $\Sigma \subseteq \mathsf{N}_\mathsf{C}(\mathcal{T}) \cup \mathsf{N}_\mathsf{R}(\mathcal{T})$. We next show that there exists a function $f_{\mathcal{T}, \Sigma} : \mathbb{N} \to \mathbb{N}$ s.t. $f_{\mathcal{T}, \Sigma}(n)$ is an upper bound on the size of extensions of predicates bounded w.r.t. $\mathcal{T}$ and Σ in the models of $(\mathcal{T}, \Sigma, \mathcal{A})$, for any ABox $\mathcal{A}$ over the signature of $\mathcal{T}$ of size n. Note that we only consider the case where bounded predicates and predicates in Σ are concept names, as roles can be eliminated using a trick similar to the one in Section 4.1.

To compute $f_{\mathcal{T}, \Sigma}$, we once again rely on integer programming. To this end, we introduce enriched linear programs. An *enriched integer linear program (enriched ILP)*, is a pair $\mathfrak{I} = (a_1 \cdot x_1 + \cdots + a_n \cdot x_n, \mathcal{S})$, where $\mathcal{S} = (V, \mathcal{E}, V_F, V_I, I)$ is an enriched system of integer linear inequalities, $x_1, \ldots, x_n \in V$ and $a_1, \ldots, a_n$ are integers. An *optimal solution* to $\mathfrak{I}$ is a solution S of $\mathcal{S}$ that maximizes $a_1 \cdot x_1 + \cdots + a_n \cdot x_n$, and the value $a_1 \cdot S(x_1) + \cdots + a_n \cdot S(x_n)$ is the *optimal value* of $\mathfrak{I}$.

We next show that for a concrete ABox $\mathcal{A}$ over the signature of $\mathcal{T}$ and a bounded concept name B, we can compute an enriched ILP whose optimal value acts as an upper bound on the size of the extension of B in the models of $(\mathcal{T}, \Sigma, \mathcal{A})$.

Proposition 4. *Let $\mathcal{T}$ be an $\mathcal{ALCHOIQ}$ TBox, $\Sigma \subseteq \mathsf{N}_\mathsf{C}(\mathcal{T})$, B be a concept name bounded w.r.t. $\mathcal{T}$ and Σ, and $\mathcal{A}$ be an ABox over the signature of $\mathcal{T}$. Then, there exists an ILP $\mathfrak{I}_{(\mathcal{T}, \Sigma, \mathcal{A})}^B = (a_1 \cdot x_1 + \cdots + a_k \cdot x_k, \mathcal{S})$ that has a finite*

optimal value $b_{\mathcal{T}, \Sigma, \mathcal{A}}^B$, and $|B^\mathcal{I}| \leq b_{\mathcal{T}, \Sigma, \mathcal{A}}^B$, for every model $\mathcal{I}$ of $(\mathcal{T}, \Sigma, \mathcal{A})$. Moreover, let $\mathcal{S} = (V, \mathcal{E}, V_F, V_I, I)$ and let c be the maximum among the absolute values of integers occurring in $\mathcal{S}$. We have $|V|, |\mathcal{E}|, |I| \leq 3(2^{n_\mathcal{T}} \cdot m_\mathcal{T})^{|\mathcal{T}| \cdot c_\mathcal{T} + 2}$, $|V_F| = |V_I| = 0$, and $c \leq \max\{1, c_\mathcal{T}, |\mathcal{A}|\}$.

Optimal solutions to enriched integer linear programs can be computed using ordinary integer programming techniques. Moreover, known integer programming results that show there is an upper bound on the optimal value of an ordinary integer program can also be transfered to enriched linear programs.

Proposition 5. *Let $\mathfrak{I} = (a_1 \cdot x_1 + \cdots + a_n \cdot x_n, \mathcal{S})$ be an enriched ILP, where $\mathcal{S} = (V, \mathcal{E}, V_F, V_I, I)$, and assume that $\mathfrak{I}$ has a finite optimal value b. Then $b \leq 4(|a_1| + 1) \cdots (|a_n| + 1) \cdot (c + 1)^{|V| \cdot (|\mathcal{E}| + |V_F| + |V_I| + |I|)}$, where c is the maximum among the absolute values of integers occurring in $\mathcal{S}$.*

Theorem 2. *Let $\mathcal{T}$ be an $\mathcal{ALCHOIQ}$ TBox, $\Sigma \subseteq \mathsf{N}_\mathsf{C}(\mathcal{T})$, and $n \geq 0$. For every concept name B bounded w.r.t $\mathcal{T}$, and Σ, every ABox $\mathcal{A}$ over the signature of $\mathcal{T}$ of size n and every model $\mathcal{I}$ of $(\mathcal{T}, \Sigma, \mathcal{A})$:*

$$|B^\mathcal{I}| \leq 4 \cdot (2^{e^{|\mathcal{T}| \cdot c_\mathcal{T}}} + 1) \cdot (c + 1)^{18 \cdot e^{(2|\mathcal{T}| \cdot c_\mathcal{T} + 4)}},$$

where $e = (2^{n_\mathcal{T}} \cdot m_\mathcal{T})$ and $c = \max\{1, c_\mathcal{T}, n\}$.

Theorem 2 shows that $f_{\mathcal{T}, \Sigma}(n) = \mathcal{O}(n^k \cdot 2^{|\mathcal{T}|^k})$, where k is a constant, i.e., f is doubly-exponential in the size of $\mathcal{T}$, but only polynomial in the size of the ABox. Note that the bound computed by $f_{\mathcal{T}, \Sigma}$ is optimal in the sense that there exists a simple TBox that creates a binary tree of exponential depth, whose every node belongs to a bounded predicate. Such a TBox is later used in the proof of Theorem 4.

Weak boundedness. Given an $\mathcal{ALCHOIQ}$ KB $\mathcal{K}$, it is not too hard to adapt the construction for fi-satisfiability to obtain an enriched system $\mathcal{S}_\mathcal{K}$ that is exponential in the size of $\mathcal{K}$ and whose solutions correspond to the models of $\mathcal{K}$. To decide whether a concept name B is bounded w.r.t. $\mathcal{K}$, we add a function that maximizes the extension of B on top of $\mathcal{S}_\mathcal{K}$. This results in an enriched ILP that has a finite optimal value if and only if B is bounded w.r.t. $\mathcal{K}$

Theorem 3. W-BOUNDEDNESS($\mathcal{ALCHOIQ}$) *is* CONEXPTIME-*complete. Furthermore, for a given KB $\mathcal{K}$, we can compute an integer bound b that is doubly-exponential in the size of $\mathcal{K}$ s.t. for all models $\mathcal{I}$ of $\mathcal{K}$, $|p^\mathcal{I}| \leq b$ holds, for all predicates p bounded w.r.t. $\mathcal{K}$.*

5 Relaxed Safety for Rule-base Queries

We discuss here how our results presented above can be applied in the context of ontology-based access to data.

Of key importance in OBDA is the problem of answering *ontology-mediated queries (OMQs)* over ABoxes, which is defined in terms of logical entailment in a DL KB. Most often an OMQ is a pair $\mathcal{Q} = (\mathcal{T}, Q)$ of a TBox $\mathcal{T}$ and a *conjunctive query (CQ)* Q. Such an OMQ can be evaluated over an ABox $\mathcal{A}$ by computing the so-called *certain answer* to Q over the KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. OMQs can be seen as database queries, but they are posed over relational structures with at most binary relations, and they are written in an expressive ontology-based query language.

Many authors have studied OMQs for various DLs, yet for $\mathcal{ALCHOIQ}$ this problem is not well-understood. In particular, for OMQs with CQs we know decidability (without an upper bound) due to [Rudolph and Glimm, 2010] and a CO2NEXPTIME lower bound due to [Glimm *et al.*, 2011]. When CQs are replaced by the more expressive *Datalog queries*, we immediately run into undecidability, even for much less expressive DLs (see, e.g., [Levy and Rousset, 1998; Rosati, 2007]). One can regain decidability by applying the *DL-safety* restriction on Datalog queries, which requires query variables to be guarded by predicates that do not occur in the TBox [Motik *et al.*, 2005; Rosati, 2005]. Unfortunately, this restricts reasoning about anonymous objects significantly. Our next goal is to show that our results can be used to relax DL-safety to support reasoning about anonymous objects, while retaining decidability. In a nutshell, the new safety condition requires all query variables to be guarded by a predicate that is bounded by the input TBox and some closed predicates. We will make this more formal next.

We assume countably infinite set N_D and N_V of *Datalog predicates* and *variables*, respectively. We assume N_D to be disjoint from N_C and N_R. Each $p \in N_D$ has a non-negative integer *arity*. Each $t \in N_V \cup N_I$ is a term. An *atom* is an expression of the form $q(t_1, \ldots, t_n)$, where $t_1, \ldots, t_n$ are terms and one of the following holds: (i) $q \in N_D$ and q is n-ary, (ii) $q \in N_C$ and $n = 1$, or (iii) $q \in N_R$ and $n = 2$. An atom $q(\vec{t})$ is *ground* if it contains no variables. A *rule* is an expression of the form $q_0(\vec{t_0}) \leftarrow q_1(\vec{t_1}), \ldots, q_m(\vec{t_m})$ such that (i) each $q_j(\vec{t_j})$, $0 \leq j \leq m$, is an atom, (ii) $q_0 \in N_D$, and (iii) for each variable X that occurs in $\vec{t_0}$, there is $1 \leq k \leq m$ s.t. X occurs in $\vec{t_k}$. A *program* is a finite set of rules.

To define the semantics, we extend interpretations to also assign meaning to the symbols in N_D. I.e. in an interpretation $\mathcal{I}$, the function $\cdot^{\mathcal{I}}$ assigns some n-ary relation $p^{\mathcal{I}} \subseteq (\Delta^{\mathcal{I}})^n$ to each n-ary predicate symbol $p \in N_D$. An *assignment* π for an interpretation $\mathcal{I}$ is any function $\pi : N_V \cup N_I \to \Delta^{\mathcal{I}}$ such that $\pi(c) = c$ for all $c \in N_I$. For a rule ρ of the form $q_0(\vec{t_0}) \leftarrow q_1(\vec{t_1}), \ldots, q_m(\vec{t_m})$, and an assignment π for $\mathcal{I}$, we write $\mathcal{I}, \pi \models \rho$ in case $\pi(\vec{t_1}) \in q_1^{\mathcal{I}}, \ldots, \pi(\vec{t_m}) \in q_m^{\mathcal{I}}$ implies $\pi(\vec{t_0}) \in q_0^{\mathcal{I}}$. An interpretation $\mathcal{I}$ is a *model* of a program P if $\mathcal{I}, \pi \models \rho$ holds for every rule $\rho \in P$ and every assignment π.

An *OMQ* is a tuple $\mathcal{Q} = (\mathcal{T}, \Sigma, P, q)$, where $\mathcal{T}$ is a TBox, $\Sigma \subseteq N_C \cup N_R$, P is a finite set of rules, and $q \in N_D$ is the *output* predicate. The *(output tuple) entailment problem* is to decide, given an ABox $\mathcal{A}$, an OMQ $\mathcal{Q}$ as above and tuple of $\vec{t}$ of individuals, whether $\vec{t} \in q^{\mathcal{I}}$ holds for every $\mathcal{I}$ that is a model of $\mathcal{K} = (\mathcal{T}, \Sigma, \mathcal{A})$ and P. If it is true, $\vec{t}$ is said to be *in the answer to $\mathcal{Q}$ over $\mathcal{A}$*.

To ensure decidability of OMQs just introduced, we define a new safety condition that exploits predicate boundedness.

Definition 4 (Safe OMQs). *A rule ρ is called* safe *for an OMQ $\mathcal{Q} = (\mathcal{T}, \Sigma, P, q)$ if every variable of ρ occurs in the body of ρ in an atom $p(\vec{t})$ such that $p \in N_D$ or p is bounded by $\mathcal{T}$ and Σ. We say $\mathcal{Q}$ is safe if every $\rho \in P$ is safe for $\mathcal{Q}$.*

Example 3. *Consider an OMQ $\mathcal{Q} = (\mathcal{T}', \{\text{Empl}\}, \{\rho\}, \text{pair})$, where $\mathcal{T}'$ is some company ontology that includes the TBox $\mathcal{T}$*

from Example 1, and ρ is the following rule:

$$\text{pair}(X, Y) \leftarrow \quad \text{Empl}(X), \text{Empl}(Y), \text{Proj}(Z)$$
$$\text{assgndTo}(X, Z), \text{assgndTo}(Y, Z)$$

Intuitively, this query computes pairs of employees working on a common project. Observe that $\mathcal{Q}$ is safe, but is not DL-safe in the sense of [Motik et al., 2005; Rosati, 2005].

We can now characterize the complexity of query answering in the proposed language.

Theorem 4. *The entailment problem for safe OMQs is* CO2NEXPTIME-*complete. The problem is* CONP-*complete in data complexity, i.e., when the size of the input TBox is assumed to be bounded by a constant.*

Proof (sketch). Assume an OMQ $\mathcal{Q} = (\mathcal{T}, \Sigma, P, q)$, an ABox $\mathcal{A}$, and a tuple $\vec{t}$ of individuals. The upper bound can be shown by a non-deterministic procedure to check that $\vec{t}$ *does not* belong to the answer to $\mathcal{Q}$ over $\mathcal{A}$. Exploiting Theorem 2, the procedure guesses a structure that extends the input ABox with up to a double exponential number of fresh objects belonging to bounded predicates. It then checks that the structure (i) is a model of the program P, (ii) it is consistent with the TBox $\mathcal{T}$, and (iii) it falsifies the atom $q(\vec{t})$. If the size of TBoxes is fixed by a constant, this structure has polynomial size. The lower bound is shown by encoding a suitable version of the *tiling problem*. The lower bound for data complexity comes from instance checking in expressive DLs [Schaerf, 1993]. □

We remark that Theorem 4 also implies (as a special case) that the data complexity of satisfiability of $\mathcal{ALCHOIQ}$ KBs with closed predicates is NP-complete.

6 Conclusion

In this paper we have presented a method to reasOAon about the amount of anonymous objects in the models of KBs written in an expressive DL with closed predicates. This provides a new tool to aid the design of ontologies, also opening the way for sophisticated yet decidable reasoning tasks for data management. One such application is answering rule-based queries over ontologies expressed in the DL $\mathcal{ALCHOIQ}$ with closed predicates, which was discussed in Section 5. Another application, which will be discussed in detail in a future paper, is in verification of temporal properties of evolving graph databases: we believe that using the key idea of Section 5 we can identify new settings with decidable verification tasks, e.g., by developing methods to identify *state bounded* systems in the sense of [Bagheri Hariri *et al.*, 2013]. Implementing our method to check boundedness is another challenging task for future work; we cannot explicitly build inequality systems and reuse existing integer programming solvers, because this would require (best-case) exponential time. A promising way to efficiently recognize predicate boundedness is to consider small systems of inequalities that provide a sound (but incomplete) approximation of the full systems defined here.

Acknowledgements

References

[Baader *et al.*, 2003] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[Bagheri Hariri *et al.*, 2013] Babak Bagheri Hariri, Diego Calvanese, Giuseppe De Giacomo, Alin Deutsch, and Marco Montali. Verification of relational data-centric dynamic systems with external services. In *Proc. of PODS 2013*. ACM, 2013.

[Bednarczyk *et al.*, 2020] Bartosz Bednarczyk, Franz Baader, and Sebastian Rudolph. Satisfiability and query answering in description logics with global and local cardinality constraints. In *Proc. of ECAI 2020*, June 2020.

[Berardi *et al.*, 2005] Daniela Berardi, Diego Calvanese, and Giuseppe De Giacomo. Reasoning on UML class diagrams. *Artificial Intelligence*, 168:70 – 118, 2005.

[Calvanese, 1996] Diego Calvanese. Finite model reasoning in description logics. In *Proc. of KR 1996*, pages 292–303. Morgan Kaufmann, 1996.

[Fan and Geerts, 2010] Wenfei Fan and Floris Geerts. Relative information completeness. *ACM Trans. Database Syst.*, 35(4):27:1–27:44, 2010.

[Franconi *et al.*, 2011] Enrico Franconi, Yazmin Angélica Ibáñez-García, and Inanç Seylan. Query answering with DBoxes is hard. *Electr. Notes Theor. Comput. Sci.*, 278:71–84, 2011.

[Glimm *et al.*, 2011] Birte Glimm, Yevgeny Kazakov, and Carsten Lutz. Status QIO: an update. In *Proc. DL 2011*, volume 745 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011.

[Gogacz *et al.*, 2020a] Tomasz Gogacz, Víctor Gutiérrez-Basulto, Yazmin Angélica Ibáñez-García, Filip Murlak, Magdalena Ortiz, and Mantas Šimkus. Ontology focusing: Knowledge-enriched databases on demand. In *Proc. of ECAI 2020*. IOS Press, 2020.

[Gogacz *et al.*, 2020b] Tomasz Gogacz, Sanja Lukumbuzya, Magdalena Ortiz, and Mantas Simkus. Datalog rewritability and data complexity of ALCHOIF with closed predicates. In Diego Calvanese, Esra Erdem, and Michael Thielscher, editors, *Proc. of KR 2020*, 2020.

[Levy and Rousset, 1998] Alon Y. Levy and Marie-Christine Rousset. Combining horn rules and description logics in CARIN. *Artif. Intell.*, 104(1-2):165–209, 1998.

[Lutz *et al.*, 2005] Carsten Lutz, Ulrike Sattler, and Lidia Tendera. The complexity of finite model reasoning in description logics. *Inf. Comput.*, 199(1-2):132–171, 2005.

[Lutz *et al.*, 2013] Carsten Lutz, Inanç Seylan, and Frank Wolter. Ontology-based data access with closed predicates is inherently intractable(sometimes). In *Proc. of IJCAI 2013*. IJCAI/AAAI, 2013.

[Motik *et al.*, 2005] Boris Motik, Ulrike Sattler, and Rudi Studer. Query answering for OWL-DL with rules. *J. Web Sem.*, 3(1):41–60, 2005.

[Pratt-Hartmann, 2005] Ian Pratt-Hartmann. Complexity of the two-variable fragment with counting quantifiers. *Journal of Logic, Language and Information*, 14(3):369–395, 2005.

[Rosati, 2005] Riccardo Rosati. On the decidability and complexity of integrating ontologies and rules. *J. Web Sem.*, 3(1):61–73, 2005.

[Rosati, 2007] Riccardo Rosati. The limits of querying ontologies. In *Proc. of ICDT 2007*. Springer, 2007.

[Rudolph and Glimm, 2010] Sebastian Rudolph and Birte Glimm. Nominals, inverses, counting, and conjunctive queries or: Why infinity is your friend! *J. Artif. Intell. Res.*, 39:429–481, 2010.

[Schaerf, 1993] Andrea Schaerf. On the complexity of the instance checking problem in concept languages with existential quantification. *J. Intell. Inf. Syst.*, 2(3):265–278, 1993.

[Schneider and Šimkus, 2020] Thomas Schneider and Mantas Šimkus. Ontologies and data management: A brief survey. *Künstliche Intell.*, 34(3):329–353, 2020.

[Xiao *et al.*, 2018] Guohui Xiao, Diego Calvanese, Roman Kontchakov, Domenico Lembo, Antonella Poggi, Riccardo Rosati, and Michael Zakharyaschev. Ontology-based data access: A survey. In Jérôme Lang, editor, *Proc. of IJCAI 2018*, pages 5511–5519. ijcai.org, 2018.