Data Complexity in Expressive Description Logics with Path Expressions

Bartosz Bednarczyk

Computational Logic Group, TU Dresden, Germany
Institute of Computer Science, University of Wrocław, Poland

Abstract

We investigate the data complexity of the satisfiability problem for the very expressive description logic $\mathcal{ZOIQ}$ (a.k.a. $\mathcal{ALCHb}^{\mathsf{Self}}_{\mathsf{reg}}\mathcal{OIQ}$) over quasi-forests and establish its NP-completeness. This completes the data complexity landscape for decidable fragments of $\mathcal{ZOIQ}$, and reproves known results on decidable fragments of OWL2 ($\mathcal{SR}$ family). Using the same technique, we establish coNEXPTIME-completeness (w.r.t. the combined complexity) of the entailment problem of rooted queries in $\mathcal{ZIQ}$.

1 Introduction

Formal ontologies are essential in nowadays approaches to safe and trustworthy symbolic AI, serving as the backbone of the Semantic Web, peer-to-peer data management, and information integration. Such applications often resort to managing and reasoning about graph-structure data. A premier choice for the reasoning framework is then the well-established family of logical formalisms known as *description logics* (DLs) [Baader *et al.*, 2017] that underpin the logical core of OWL 2 by the W3C [Hitzler *et al.*, 2012]. Among the plethora of various features available in extensions of the basic DL called $\mathcal{ALC}$, an especially prominent one is $\cdot_{\mathsf{reg}}$, supported by the popular $\mathcal{Z}$-family of DLs [Calvanese *et al.*, 2009]. Using $\cdot_{\mathsf{reg}}$, one can specify regular path constraints and hence, allow the user to navigate the underlying graph data-structure. In recent years, many extensions of $\mathcal{ALC}_{\mathsf{reg}}$ for ontology-engineering were proposed [Bienvenu *et al.*, 2014; Calvanese *et al.*, 2016; Ortiz and Simkus, 2012], and the complexity of their reasoning problems is mostly well-understood: Calvanese *et al.* [2009; 2016]; Bednarczyk *et al.* [2019; 2022].

Vardi [1982] observed that measuring the user's data and the background ontology equally is not realistic, as the data tends to be huge in comparison to the ontology. This gave rise to the notion of *data complexity*: the ontology (TBox) is fixed upfront and only the user's data (ABox) varies. The *satisfiability* problem for DLs is usually NP-complete w.r.t. the data complexity, including the two-variable counting logic [Pratt-Hartmann, 2009] (encoding DLs up to $\mathcal{ALCBIOQ}\mathsf{Self}$) as well as $\mathcal{SROIQ}$ [Kazakov, 2008], the logical core of OWL2. Regarding DLs with path expressions, NP-completeness of $\mathcal{ALCI}^{\mathsf{Self}}_{\mathsf{reg}}$ [Jung *et al.*, 2018] was established only recently.

1.1 Our Main Contribution and a Proof Overview

We study the data complexity of the satisfiability problem for decidable sublogics of $\mathcal{ZOIQ}$ (a.k.a. $\mathcal{ALCHb}^{\mathsf{Self}}_{\mathsf{reg}}\mathcal{OIQ}$), and establish NP-completeness of $\mathcal{ZIQ}$, $\mathcal{ZOQ}$, and $\mathcal{ZOI}$. As all the mentioned DLs possess the *quasi-forest model property* (*i.e.* every satisfiable knowledge-base (KB) has a forest-like model), for the uniformity of our approach we focus on the satisfiability of $\mathcal{ZOIQ}$ over quasi-forests. Calvanese et al. [2009] proved that quasi-forest-satisfiability of $\mathcal{ZOIQ}$ is EXPTIME-complete w.r.t. the combined complexity. Unfortunately, their approach is automata-based and relies on an internalisation of ABoxes inside $\mathcal{ZOIQ}$-concepts, and thus cannot be used to infer tight bounds w.r.t. the data complexity.

We employ the aforementioned algorithm of Calvanese et al. as a black box and design a novel algorithm for quasi-forest-satisfiability of $\mathcal{ZOIQ}$-KBs. In our approach we construct a quasi-forest model in two steps, *i.e.* we construct its root part (the *clearing*) separately from its subtrees. Our algorithm first pre-computes (an exponential w.r.t. the size of the TBox but of constant size if the TBox is fixed) set of quasi-forest-satisfiable $\mathcal{ZOIQ}$-concepts that indicate possible subtrees that can be "plugged in" to the clearing of the intended model. Then it guesses (in NP) the intended clearing and verifies its consistency in PTIME based on the pre-computed concepts and roles. For the feasibility of our "modular construction" a lot of bookkeeping needs to be done. Most importantly, certain *decorations* are employed to decide the satisfaction of *automata concepts* and *number restrictions* by elements in an incomplete and fragmented forest. The first type of decorations, given an automaton $\mathcal{A}$, aggregates information about existing paths realising $\mathcal{A}$ and starting at one of the roots of the intended model. As a single such path may visit several subtrees, we cut such paths into relevant pieces and summarise them by means of "shortcut" roles and $\mathcal{ZOIQ}$-concepts describing paths fully contained inside a single subtree. The second type of decorations "localise" counting in the presence of nominals, as the nominals may have successors outside their own subtree and the clearing. These two "small tricks", obfuscated by various technical difficulties, are the core ideas underlining our proof. We conclude the paper by presenting how our algorithm can be adapted to the entailment problem of rooted queries over $\mathcal{ZIQ}$. The key idea here is to guess an "initial segment" of a quasi-forest with no query match, and check if it can be extended to a full model of the input KB.

2 Preliminaries

We assume familiarity with description logics (DLs) [Baader *et al.*, 2017], formal languages, and complexity [Sipser, 2013]. We fix countably-infinite pairwise-disjoint sets $\mathbf{N_I}, \mathbf{N_C}, \mathbf{N_R}$ of *individual*, *concept*, and *role names*. An *interpretation* $\mathcal{I} := (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty domain $\Delta^{\mathcal{I}}$ and a function $\cdot^{\mathcal{I}}$ that maps $\mathsf{a} \in \mathbf{N_I}$ to $\mathsf{a}^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, $\mathrm{A} \in \mathbf{N_C}$ to $\mathrm{A}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, and $r \in \mathbf{N_R}$ to $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The set of *simple roles* $\mathbf{N}_{\mathbf{R}}^{\mathrm{sp}}$ is defined with the grammar $s ::= r \in \mathbf{N_R} \mid s^- \mid s{\cap}s \mid s{\cup}s \mid s{\setminus}s$. Simple roles are interpreted as expected by invoking the underlying set-theoretic operations, *e.g.* $\mathcal{I}$ interprets $(r \cup (s \setminus t))^-$ as the inverse of $(r^{\mathcal{I}} \cup (s^{\mathcal{I}} \setminus t^{\mathcal{I}}))$. The superscripts * and $^+$ denote the Kleene's star and plus, while $\cdot$ denotes concatenation. A *path* in $\mathcal{I}$ is a word $\rho := \rho_1 \ldots \rho_{|\rho|}$ in $(\Delta^{\mathcal{I}})^+$ such that for all $i < |\rho|$ there is a role name $r \in \mathbf{N}_{\mathbf{R}}^{\mathrm{sp}}$ with $(\rho_i, \rho_{i+1}) \in r^{\mathcal{I}}$.

Automata. We employ standard notion of nondeterministic automata (NFAs). For states q and q' of an NFA $\mathcal{A}$, $\mathcal{A}_{\mathsf{q},\mathsf{q}'}$ denotes the corresponding NFA with the initial (resp. final) state switched to q (resp. q'). NFAs use roles from $\mathbf{N}_{\mathbf{R}}^{\mathrm{sp}}$ and tests C? for concepts C as alphabet. A path ρ *realises* $\mathcal{A}$ ($\rho \models \mathcal{A}$) if $\mathcal{A}$ accepts some word $w_1 r_1 \ldots w_{|\rho|-1} r_{|\rho|-1} w_{|\rho|}$, where $r_i \in \mathbf{N}_{\mathbf{R}}^{\mathrm{sp}}$ and w_i are (possibly empty) sequences of tests, satisfying $(\rho_i, \rho_{i+1}) \in r_i^{\mathcal{I}}$ and $\rho_i \in \mathrm{C}^{\mathcal{I}}$ for all $i \leq |\rho|$ and tests C? in w_i.

Expressive DLs. The DL $\mathcal{ZOIQ}$ [Calvanese *et al.*, 2009] is the main object of our study. For all $\mathrm{A} \in \mathbf{N_C}$, $s \in \mathbf{N}_{\mathbf{R}}^{\mathrm{sp}}$, $\circ \in \mathbf{N_I}$, $n \in \mathbb{N}$ and NFA $\mathcal{A}$, we define $\mathcal{ZOIQ}$-concepts C with:

$$\bot \mid \mathrm{A} \mid \{\circ\} \mid \neg\mathrm{C} \mid \mathrm{C} \sqcap \mathrm{C} \mid \exists s.\mathrm{C} \mid ({\geq} n\ s).\mathrm{C} \mid \exists s.\mathsf{Self} \mid \exists \mathcal{A}.\mathrm{C}$$

We denote $\top := \neg\bot$, $\mathrm{C} \sqcup \mathrm{C}' := \neg(\neg\mathrm{C} \sqcap \neg\mathrm{C}')$, $({\leq} n\ s).\mathrm{C} := \neg({\geq} n{+}1\ s).\neg\mathrm{C}$, $({=} n\ s).\mathrm{C} := ({\leq} n\ s).\mathrm{C} \sqcap ({\geq} n\ s).\mathrm{C}$, and $\forall\heartsuit.\mathrm{C} := \neg\exists\heartsuit.\neg\mathrm{C}$, where $\heartsuit$ is a simple role or an NFA.

Name	Syntax	Semantics: $d \in \mathrm{C}^{\mathcal{I}}$ if
top/bottom	$\top / \bot$	always/never
nominal	$\{\circ\}$	$\mathrm{d} = \circ^{\mathcal{I}}$
negation	$\neg\mathrm{C}$	$\mathrm{d} \notin \mathrm{C}^{\mathcal{I}}$
intersection	$\mathrm{C}_1 \sqcap \mathrm{C}_2$	$\mathrm{d} \in \mathrm{C}_1^{\mathcal{I}}$ and $\mathrm{d} \in \mathrm{C}_2^{\mathcal{I}}$
existential restriction	$\exists s.\mathrm{C}$	$\exists \mathrm{e}\left(\mathrm{e} \in \mathrm{C}^{\mathcal{I}} \wedge (\mathrm{d}, \mathrm{e}) \in s^{\mathcal{I}}\right)$
number restriction	$({\geq} n\ s).\mathrm{C}$	$\lvert\{\mathrm{e} \in \mathrm{C}^{\mathcal{I}} \mid (\mathrm{d}, \mathrm{e}){\in}s^{\mathcal{I}}\}\rvert \geq n$
Self concept	$\exists s.\mathsf{Self}$	$(\mathrm{d}, \mathrm{d}) \in s^{\mathcal{I}}$
automaton concept	$\exists \mathcal{A}.\mathrm{C}$	$\exists$path $\mathrm{d}{\cdot}\rho{\cdot}\mathrm{e} \models \mathcal{A}$ & $\mathrm{e} \in \mathrm{C}^{\mathcal{I}}$

We define the DLs $\mathcal{ZOI}$, $\mathcal{ZIQ}$, and $\mathcal{ZOQ}$, respectively, by dropping (a) number restrictions, (b) nominals, and (c) role inverses from the syntax of $\mathcal{ZOIQ}$. W.l.o.g. we allow for regular expression (with tests) in place of NFAs in above syntax.

A $\mathcal{ZOIQ}$*-KB* $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ is composed of two sets of *axioms*, an *ABox* $\mathcal{A}$ and a *TBox* $\mathcal{T}$. An interpretation $\mathcal{I}$ is a *model* of $\mathcal{K}$ ($\mathcal{I} \models \mathcal{K}$) if it satisfies all the axioms of $\mathcal{K}$. The content of ABoxes and TBoxes, as well as the notion of their satisfaction is defined below, assuming that $\mathsf{a}, \mathsf{b} \in \mathbf{N_I}$, $\mathrm{A} \in \mathbf{N_C}$, $r \in \mathbf{N_R}$, $s, t \in \mathbf{N}_{\mathbf{R}}^{\mathrm{sp}}$, and C, D are $\mathcal{ZOIQ}$-concepts.

Axiom α	$\mathcal{I} \models \alpha$, if	
$\mathrm{C} \sqsubseteq \mathrm{D}$, $s \subseteq t$	$\mathrm{C}^{\mathcal{I}} \subseteq \mathrm{D}^{\mathcal{I}}$, $s^{\mathcal{I}} \subseteq t^{\mathcal{I}}$	TBox $\mathcal{T}$
$\mathrm{A}(\mathsf{a})$, $r(\mathsf{a},\mathsf{b})$ $\mathsf{a} \approx \mathsf{b}$, $\neg\alpha$	$\mathsf{a}^{\mathcal{I}} \in \mathrm{A}^{\mathcal{I}}$, $(\mathsf{a}^{\mathcal{I}}, \mathsf{b}^{\mathcal{I}}) \in r^{\mathcal{I}}$ $\mathsf{a}^{\mathcal{I}} = \mathsf{b}^{\mathcal{I}}$, $\mathcal{I} \not\models \alpha$	ABox $\mathcal{A}$

Note that ABoxes do not involve complex concepts. Otherwise, data complexity of $\mathcal{ALC}_{\mathsf{reg}}$ is already ExpTime-hard.

A TBox is in *Scott's normal form* if its axioms have the form:

$$\mathrm{A} \equiv \mathrm{B}, \quad \mathrm{A} \equiv \neg\mathrm{B}, \quad \mathrm{A} \equiv \{\circ\}, \quad \mathrm{A} \equiv \mathrm{B} \sqcap \mathrm{B}', \quad r = s,$$
$$\mathrm{A} \equiv \exists \mathcal{A}_{\mathsf{q},\mathsf{q}'}.\top, \quad \mathrm{A} \equiv \exists r.\mathsf{Self}, \quad \mathrm{A} \equiv ({\geq} n\ r).\top$$

for $\mathrm{A}, \mathrm{B}, \mathrm{B}' \in \mathbf{N_C} \cup \{\top, \bot\}$, $\circ \in \mathbf{N_I}$, $r \in \mathbf{N_R}$, $s \in \mathbf{N}_{\mathbf{R}}^{\mathrm{sp}}$, $n \in \mathbb{N}$, and NFAs $\mathcal{A}_{\mathsf{q},\mathsf{q}'}$. Here $\mathcal{I} \models \mathrm{C} \equiv \mathrm{D}$ iff $\mathrm{C}^{\mathcal{I}} = \mathrm{D}^{\mathcal{I}}$. W.l.o.g. we focus only on KBs with TBoxes as above [Appendix B].

Queries. We focus on *conjunctive queries* (CQs) defined by the grammar: $q ::= \mathrm{A}(x) \mid r(x, y) \mid q \wedge q$, where $\mathrm{A} \in \mathbf{N_C}$, $r \in \mathbf{N_R}$, and x, y being either variables or individual names. An interpretation $\mathcal{I}$ *satisfies* a query q ($\mathcal{I} \models q$), if there is an assignment η (a *match*, defined as expected), mapping variables to domain elements and individual names $\mathsf{a} \in \mathbf{N_I}$ to the corresponding $\mathsf{a}^{\mathcal{I}}$, under which q evaluates to true. A CQ is *rooted* if it has at least one individual name and its query graph is connected. A KB *entails* q if every of its models satisfies q.

Problems. Let $\mathcal{DL}$ be a logic, and $\mathcal{Q}$ be a class of queries. In the *satisfiability problem* we ask if a given $\mathcal{DL}$-KB $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ has a model. In the *(rooted) query entailment problem* we ask if a given $\mathcal{DL}$-KB $\mathcal{K}$ entails a (rooted) query $q \in \mathcal{Q}$. We distinguish two ways of measuring the size of the input, which gives rise to *combined* and *data complexity*. In the first case all the components $\mathcal{A}, \mathcal{T}$ and q equally contribute to the size of the input. In the second case, $\mathcal{T}$ and q are assumed to be fixed beforehand, and thus their size is treated as constant.

Forests. We adapt the notion of *quasi-forests* [Calvanese *et al.*, 2009] under the standard set-theoretic reconstruction of the notion of an $\mathbb{N}$-forest as a prefix-closed subset of $\mathbb{N}^+$ without ε. Our notion is stricter than the original definition by Calvanese *et al.* but the differences are negligible [Appendix C].

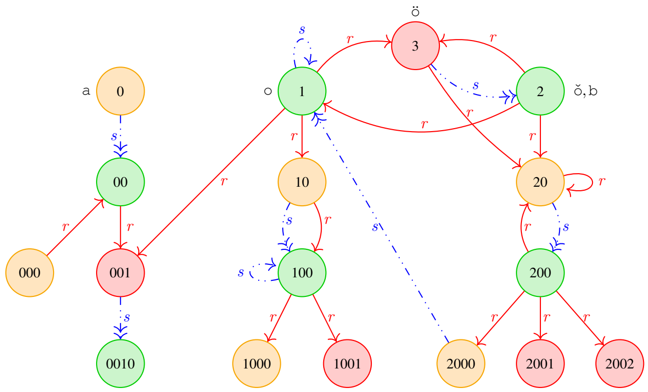
Definition 1. *Let $\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}$ and $\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}$ be finite subsets of $\mathbf{N_I}$, $\mathrm{Root} \in \mathbf{N_C}$, and $\mathit{child}, \mathit{edge}, \mathit{id} \in \mathbf{N_R}$. An interpretation $\mathcal{I}$ is an* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$***-quasi-forest*** *if its domain $\Delta^{\mathcal{I}}$ is an $\mathbb{N}$-forest,*

$$\mathrm{Root}^{\mathcal{I}} = \Delta^{\mathcal{I}} \cap \mathbb{N} = \{a^{\mathcal{I}} \mid a {\in} \mathbf{N_I}\} = \{a^{\mathcal{I}} \mid a {\in} (\mathbf{N}_{\mathbf{I}}^{\mathcal{A}} {\cup} \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})\},$$
$$\mathit{child}^{\mathcal{I}} = \{(\mathrm{d}, \mathrm{d}{\cdot}n) \mid \mathrm{d}, \mathrm{d}{\cdot}n \in \Delta^{\mathcal{I}}, n \in \mathbb{N}\},$$
$$\mathit{edge}^{\mathcal{I}} = \textstyle\bigcup_{r \in \mathbf{N_R}} r^{\mathcal{I}} \cup (r^-)^{\mathcal{I}},$$
$$\mathit{id}^{\mathcal{I}} = \{(\mathrm{d}, \mathrm{d}) \mid \mathrm{d} \in \Delta^{\mathcal{I}}\},$$

and for all roles $r^{\mathcal{I}}$ and all pairs $(\mathrm{d}, \mathrm{e}) \in r^{\mathcal{I}}$ at least one of the conditions hold: (i) both d *and* e *belong to* $\mathrm{Root}^{\mathcal{I}}$*, (ii) one of* d *or* e *is equal to $\circ^{\mathcal{I}}$ for some name $\circ \in \mathbf{N}_{\mathbf{I}}^{\mathcal{T}}$, (iii) $(\mathrm{d}, \mathrm{e}) \in \mathit{id}^{\mathcal{I}} \cup \mathit{child}^{\mathcal{I}} \cup (\mathit{child}^-)^{\mathcal{I}}$. For convenience, we refer to pairs $(\mathrm{d}, \mathrm{e}) \in r^{\mathcal{I}}$ satisfying the second condition as* backlinks*, and the ones satisfying $(\mathrm{d}, \mathrm{e}) \in \mathit{id}^{\mathcal{I}}$ as* self-loops*. The* clearing *of $\mathcal{I}$ is the restriction of $\mathcal{I}$ to* $\mathrm{Root}^{\mathcal{I}}$*. A quasi-forest is* N-bounded *if every domain element has at most* N *child-successors.* ◀

The names from $\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}$ are dubbed *nominals*, and will be usually denoted with decorated letters $\circ$. Their interpretations are usually referred as *nominal roots*. Throughout the paper we employ suitable notions from graph theory such as *node*, *root*, *child*, *parent*, or *descendant*, defined as expected in accordance with $\Delta^{\mathcal{I}}$, $\mathit{child}^{\mathcal{I}}$ and $\mathrm{Root}^{\mathcal{I}}$. For instance, d is a descendant of c whenever $(\mathrm{c}, \mathrm{d}) \in (\mathit{child}^{\mathcal{I}})^+$ holds. Consult Example 1.

Example 1. *An example* 3*-bounded* $(\{a, b\}, \{o, \ddot{o}, \breve{o}\})$*-quasi-forest* $\mathcal{I}$ *is depicted below. For readability we have omitted the interpretations of* Root, *child, id, and edge.*



The element 1 *has a unique child* 10*. The roots of* $\mathcal{I}$ *are* 0, 1, 2, *and* 3*. The nominal roots are* 1,2, *and* 3*. The pairs* (1, 001), (2000, 1), *and* (3, 20) *are example backlinks.* ◀

A *quasi-forest model* of a $\mathcal{ZOIQ}$-KB $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ is an $(\mathsf{ind}(\mathcal{A}), \mathsf{ind}(\mathcal{T}))$-quasi-forest satisfying $\mathcal{K}$, where $\mathsf{ind}(\mathcal{A})$ and $\mathsf{ind}(\mathcal{T})$ are sets of all individual names from $\mathcal{A}$ and $\mathcal{T}$. The results by Calvanese *et al.* [2009, Prop. 3.3] and Ortiz [2010, L. 3.4.1, Thm. 3.4.2] advocate the use of quasi-forest models.

Lemma 1. *Let* $\mathcal{K}{:=}(\mathcal{A}, \mathcal{T})$ *be a KB of* $\mathcal{ZOQ}$*,* $\mathcal{ZOI}$*, or* $\mathcal{ZIQ}$*. Then is an integer* N *(exponential w.r.t.* $|\mathcal{T}|$*) such that: (I)* $\mathcal{K}$ *is satisfiable iff* $\mathcal{K}$ *has an* N*-bounded quasi-forest model, and (II) for all (unions of) CQs* q*, we have* $\mathcal{K} \not\models q$ *iff there exists an* N*-bounded quasi-forest model of* $\mathcal{K}$ *violating* q*.* ◀

The quasi-forest (counter)models of $\mathcal{ZOIQ}$-KBs are *canonical* if they are N-bounded for the N guaranteed by Lemma 1. Hence, for the deciding the satisfiability and query entailment over $\mathcal{ZIQ}$, $\mathcal{ZOQ}$, and $\mathcal{ZOI}$, only the class of canonical quasi-forest models is relevant. We say that a $\mathcal{ZOIQ}$-KB is *quasi-forest satisfiable* whenever it has a canonical quasi-forest model. The quasi-forest satisfiability problem is defined accordingly. We have [Ortiz, 2010, L 3.4.1, Thm. 3.4.2] that:

Lemma 2. *The quasi-forest satisfiability problem for* $\mathcal{ZOIQ}$*-KBs is* ExpTime*-complete (w.r.t. the combined complexity). In particular, the satisfiability of* $\mathcal{ZIQ}$*,* $\mathcal{ZOQ}$*, and* $\mathcal{ZOI}$*-KBs is* ExpTime*-complete (w.r.t. the combined complexity).* ◀

3 Basic Paths in Quasi-Forests

The main goal of this section is to provide a characterisation of how paths in quasi-forests look like and how they can be decomposed into interesting pieces. A path ρ in $\mathcal{I}$ is *nameless* if it does not contain any *named* elements (*i.e.* no element of ρ has the form $\mathrm{a}^{\mathcal{I}}$ for some $\mathrm{a} \in \mathbf{N_I}$). Similarly, ρ is called an *a-subtree path* if all the members of ρ are descendants of $\mathrm{a}^{\mathcal{I}}$. In quasi-forests, nameless paths are precisely subtree paths.

Definition 2. *Let* $\mathcal{I}$ *be an* $(\mathbf{N_I^{\mathcal{A}}}, \mathbf{N_I^{\mathcal{T}}})$*-quasi-forest,* $a, b \in (\mathbf{N_I^{\mathcal{A}}} \cup \mathbf{N_I^{\mathcal{T}}})$*,* $o, \ddot{o} \in \mathbf{N_I^{\mathcal{T}}}$*, and* ρ *be a path in* $\mathcal{I}$*. We call* ρ*:*

- ***(a, b)-direct*** *if* $\rho = a^{\mathcal{I}} {\cdot} b^{\mathcal{I}}$*.*
- ***a-inner*** *if* $\rho = a^{\mathcal{I}} {\cdot} \bar{\rho}$ *for some a-subtree path* $\bar{\rho}$*.*
- ***a-roundtrip*** *if* $\rho = a^{\mathcal{I}} {\cdot} \bar{\rho} {\cdot} a^{\mathcal{I}}$ *for some a-subtree path* $\bar{\rho}$*.*
- ***(a, o)-inout*** *if* $\rho = \bar{\rho} {\cdot} o^{\mathcal{I}}$ *for some a-inner path* $\bar{\rho}$*.*
- ***(a, o)-outin*** *if the reverse of* ρ *is* (a, o)*-inout.*
- ***(a, o)-inner*** *if* $\rho = o^{\mathcal{I}} {\cdot} \bar{\rho}$ *for some a-subtree path* $\bar{\rho}$*.*
- ***(a, o, ö)-bypass*** *if* $\rho = o^{\mathcal{I}} {\cdot} \bar{\rho} {\cdot} \ddot{o}^{\mathcal{I}}$ *for an a-subtree path* $\bar{\rho}$*.*

If ρ *falls into one of the above seven categories, we call* ρ ***basic****. Paths* $\rho := \bar{\rho}{\cdot}\mathrm{d}$ *for a nameless* $\bar{\rho}$ *and a root* d *are called* ***outer****. Finally,* ρ *is* ***decomposable*** *whenever there exists a growing sequence of indices* $i_1 < i_2 < \ldots < i_k$ *with* $i_1{=}1$ *and* $i_k{=}|\rho|$ *such that for all* $j < k$ *the path* $\rho_{i_j} \ldots \rho_{i_{(j+1)}}$ *is basic.* ◀

By a careful analysis of Definitions 1–2 we can show:

Observation 1. *Take* $\mathcal{I}$*,* a*,* b*,* ρ *as in Definition 2. If* ρ *has the form* $a^{\mathcal{I}} {\cdot} \bar{\rho}$ *or* $a^{\mathcal{I}} {\cdot} \bar{\rho} {\cdot} b^{\mathcal{I}}$ *for a nameless* $\bar{\rho}$ *then* ρ *is basic.* ◀

We can now employ Observation 1 and invoke the induction over the total number of named elements in ρ to establish:

Lemma 3. *If* $\mathcal{I}$ *is a quasi-forest then every path starting from a named element is decomposable.* ◀

Basic paths are expressible in $\mathcal{ZOIQ}$ by means of regular expressions with tests. Their construction involves nominal tests $\{\mathrm{a}\}?$, "descendant of a" tests $\exists(child^{-})^{+}.\{\mathrm{a}\}?$, and role names $edge$, $child$. For instance, all $(\mathrm{a}, \mathrm{o}, \ddot{\mathrm{o}})$-bypasses match:

$$\{\mathrm{o}\}?\ edge\ \left([\exists(child^{-})^{+}.\{\mathrm{a}\}]?\ edge\right)^{+}\ \{\ddot{\mathrm{o}}\}?$$

Lemma 4. *Let* τ *be a "category" of paths from Def. 2 (including nameless and outer paths). There is an NFA* $\mathcal{A}_\tau$ *(using testing involving only individual names mentioned in* τ*) that realises for all quasi-forests* $\mathcal{I}$ *exactly the paths of the category* τ*. Moreover, for any regular language* $\mathcal{L}$ *(given as an NFA* $\mathcal{B}$ *or a reg-exp* $\mathcal{R}$*) there exists an NFA* $(\mathcal{B} \bowtie \mathcal{A}_\tau)$ *(resp.* $(\mathcal{R} \bowtie \mathcal{A}_\tau)$*) of size polynomial in* $|\mathcal{B}|$ *(resp.* $|\mathcal{R}|$*) that realises precisely the paths in* $\mathcal{I}$ *of category* τ *realising* $\mathcal{B}$ *(resp.* $\mathcal{R}$*).* ◀

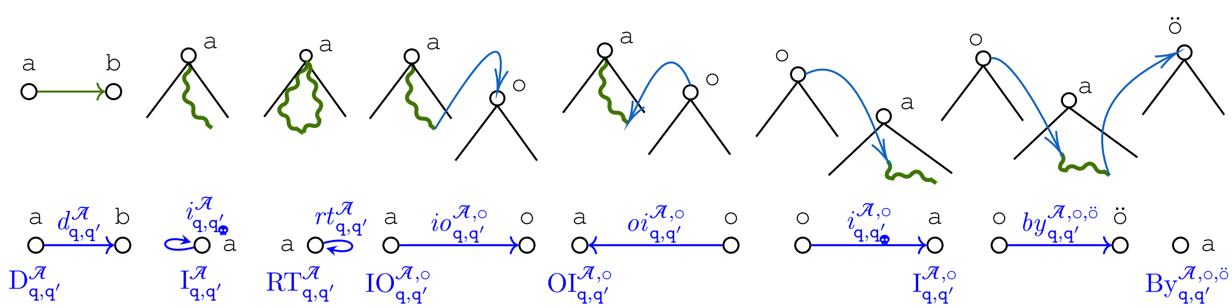


Figure 1: Basic paths and their corresponding decorations in quasi-forests, given in order of their introduction in Def. 2, 3, and 4.

4 Automata Decorations

Overview. We use Lem. 3 to devise a method of decorating the clearing of $\mathcal{I}$ with extra information about the basic paths realising an NFA $\mathcal{A}$. This helps to decide the satisfaction of concepts $\exists\mathcal{A}.\top$ in quasi-forests in a modular way, independently for the clearing of a quasi-forest and its subtrees. Note:

Fact 1. *Let* $\mathrm{d} \in (\exists\mathcal{A}.\top)^{\mathcal{I}}$ *for an NFA* $\mathcal{A}$ *and an* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*-quasi-forest* $\mathcal{I}$*. There is an* $\mathcal{A}$*-path* ρ *starting from* d *s.t.: (a)* d *is a root, (b)* ρ *is nameless, or (c)* $\rho = \bar{\rho}\cdot\hat{\rho}$ *for some outer* $\bar{\rho}$*.* ◀

Given an NFA $\mathcal{A}$ with the state set $\mathtt{Q}$, by the ***$\mathcal{A}$* reachability concepts** $\mathbf{C}_{\rightsquigarrow}(\mathcal{A})$ we mean the set $\{\mathrm{Reach}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'} \mid \mathtt{q},\mathtt{q}' \in \mathtt{Q}\}$. Ideally, $\mathrm{Reach}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$ should label precisely the roots of quasi-forests satisfying $\exists\mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top$. With Lemma 5 we explain the desired "modularity" condition. It simply says that if $\mathrm{Reach}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$-concepts are interpreted as desired, then the verification of whether an element d satisfies $\exists\mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top$ boils to (i) testing whether d is labelled with $\mathrm{Reach}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$ if d is a root, or (ii) testing existence of a certain path which fully contained in the subtree of d, possibly except the last element which can also be the root of d or a nominal. We rely on the NFAs $\mathcal{A}_{\mathrm{nmls}}$ and $\mathcal{A}_{\mathrm{outr}}$ from Lemma 4 that detect nameless and outer paths.

Lemma 5. *Let* $\mathcal{A}$ *be an NFA with the set of states* $\mathtt{Q}$*, and let* $\mathcal{I}$ *be an* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*-quasi-forest that interprets all* $\mathrm{Reach}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$*-concepts from* $\mathbf{C}_{\rightsquigarrow}(\mathcal{A})$ *equally to* $\mathrm{Root} \sqcap \exists\mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top$*. Then for all states* $\mathtt{q},\mathtt{q}' \in \mathtt{Q}$*, the concept* $\exists\mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top$ *is interpreted in* $\mathcal{I}$ *equally to the union of concepts:*
- $\mathrm{Root} \sqcap \mathrm{Reach}_{\mathtt{q},\mathtt{q}'}$, • $\neg\mathrm{Root} \sqcap \exists(\mathcal{A}_{\mathtt{q},\mathtt{q}'} \bowtie \mathcal{A}_{nmls}).\top$, *and*
- $\neg\mathrm{Root} \sqcap \bigsqcup_{\hat{\mathtt{q}} \in \mathtt{Q}} \exists(\mathcal{A}_{\mathtt{q},\hat{\mathtt{q}}} \bowtie \mathcal{A}_{outr}).\,(\mathrm{Root} \sqcap \mathrm{Reach}_{\hat{\mathtt{q}},\mathtt{q}'})$. ◀

Let $\mathrm{rel}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A}_{\mathtt{q},\mathtt{q}'})$ denote the above concept union, called the **$\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}$-relativisation** of $\exists\mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top$. We say that $\mathrm{d} \in \Delta^{\mathcal{I}}$ **virtually satisfies** $\exists\mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top$ whenever d satisfies $\mathrm{rel}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A}_{\mathtt{q},\mathtt{q}'})$. Thus, Lemma 5 tells us that under some extra assumptions the notion of satisfaction and virtual satisfaction coincide.

Decorations. In what follows, we define $\mathcal{ZOIQ}$-concepts and roles, dubbed *automata decorations*, intended to guarantee the desired interpretation of the reachability concepts.

Definition 3. *Given an NFA* $\mathcal{A}$ *with the state-set* $\mathtt{Q}$*, the set of* ***$(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$-concepts*** $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *is composed of the following concept names (for all states* $\mathtt{q},\mathtt{q}' \in \mathtt{Q}$ *and names* $\mathtt{o}, \ddot{\mathtt{o}} \in \mathbf{N}_{\mathbf{I}}^{\mathcal{T}}$*):*

$$\mathrm{D}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'},\ \mathrm{I}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'},\ \mathrm{RT}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'},\ \mathrm{IO}^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'},\ \mathrm{OI}^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'},\ \mathrm{I}^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'},\ \mathrm{By}^{\mathcal{A},\mathtt{o},\ddot{\mathtt{o}}}_{\mathtt{q},\mathtt{q}'}.$$

A quasi-forest $\mathcal{I}$ ***properly interprets*** $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *if* $\mathrm{C}^{\mathcal{I}} = \{\mathtt{a}^{\mathcal{I}} \mid \mathtt{a} \in \mathbf{N}_{\mathbf{I}}, \mathrm{cond}(\mathrm{C}, \mathtt{a})\}$ *for concepts and conditions as below.*

conc. C	$\mathrm{cond}(\mathrm{C}, \mathtt{a})$: "There is a path $\rho \models \mathcal{A}_{\mathtt{q},\mathtt{q}'}$ s.t.
$\mathrm{D}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$	ρ is $(\mathtt{a},\mathtt{b})$-direct for some $\mathtt{b}^{\mathcal{I}}$".
$\mathrm{I}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$	ρ is $\mathtt{a}$-inner".
$\mathrm{RT}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$	ρ is an $\mathtt{a}$-roundtrip".
$\mathrm{IO}^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'}$	ρ is an $(\mathtt{a},\mathtt{o})$-inout".
$\mathrm{OI}^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'}$	ρ is an $(\mathtt{a},\mathtt{o})$-outin".
$\mathrm{I}^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'}$	ρ is $(\mathtt{a},\mathtt{o})$-inner".
$\mathrm{By}^{\mathcal{A},\mathtt{o},\ddot{\mathtt{o}}}_{\mathtt{q},\mathtt{q}'}$	ρ is an $(\mathtt{a},\mathtt{o},\ddot{\mathtt{o}})$-bypass".

The size of $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *is clearly polynomial in* $|\mathcal{A}|\cdot|\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}|$*.* ◀

The concepts from $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ "bookkeep" the information about relevant basic paths realising NFA starting from a given root. Throughout the paper, we employ a *fresh* individual name $\ghost$ (the *ghost variable*), to stress that the constructed concepts are independent from $\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}$. With $\mathrm{C}[\ghost/\mathtt{a}]$ we denote the result of substituting $\mathtt{a}$ for all occurrences of $\ghost$ in C. Relying on NFAs from Lemma 4, we construct $\mathcal{ZOIQ}$-concepts that describe the intended behaviour of $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$-concepts. For instance, $\mathrm{RT}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$ is defined as $\exists(\mathcal{A}_{\mathtt{q},\mathtt{q}'} \bowtie \mathcal{A}_{\ghost\text{-roundtrip}}).\top$.

Lemma 6. *For all* C *from* $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$*, there exists a* $\mathcal{ZOIQ}$*-concept* $\mathrm{desc}(\mathrm{C})$ *(of size polynomial w.r.t.* $|\mathcal{A}|\cdot|\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}|$*) that uses only individual names from* $\mathbf{N}_{\mathbf{I}}^{\mathcal{T}} \cup \{\ghost\}$ *with the property that "for all* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*-quasi-forests* $\mathcal{I}$ *and* $\mathtt{a} \in \mathbf{N}_{\mathbf{I}}$ *we have:* $\mathrm{cond}(\mathrm{C}, \mathtt{a})$ *is satisfied in* $\mathcal{I}$ *iff* $\mathtt{a}^{\mathcal{I}}$ *is in* $(\mathrm{desc}(\mathrm{C})[\ghost/\mathtt{a}])^{\mathcal{I}}$*."* ◀

We next invoke Lemma 6 to rephrase the notion of proper interpretation of $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ in the language of satisfaction of $\mathcal{ZOIQ}$-concepts by the clearings of quasi-forests. We have:

Fact 2. *Let* $\mathrm{comdsc}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A}) \coloneqq \bigsqcap_{\mathrm{C} \in \mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})} (\mathrm{C} \leftrightarrow \mathrm{desc}(\mathrm{C}))$ *An* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*-quasi-forest* $\mathcal{I}$ *properly interprets* $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *iff* $\mathtt{a}^{\mathcal{I}} \in \mathrm{comdsc}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})\,[\ghost/\mathtt{a}]^{\mathcal{I}}$ *for all* $\mathtt{a} \in (\mathbf{N}_{\mathbf{I}}^{\mathcal{A}} \cup \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*.* ◀

The information provided by the proper interpretation of $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$-concepts is not sufficient to ensure the proper interpretation of the reachability concepts by the clearings of quasi-forests. The main reason is that the concepts from $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ concern only about the basic paths. As some automata constraints cannot be satisfied by basic paths, *e.g.* $\exists(\{\mathtt{o}\}?\mathit{edge}\{\ddot{\mathtt{o}}\}?\mathit{edge}\{\check{\mathtt{o}}\}?).\top$, more work needs to be done. To be able to compose basic paths into bigger pieces, another "layer of decoration" is needed: this time with fresh roles representing the endpoints of basic paths from Definition 2, as summarised by Figure 1. For instance, whenever there is an $(\mathtt{a},\mathtt{o})$-inout ρ realising $\mathcal{A}_{\mathtt{q},\mathtt{q}'}$ in a quasi-forest $\mathcal{I}$, the roots $\mathtt{a}^{\mathcal{I}}$ and $\mathtt{o}^{\mathcal{I}}$ are going to be linked by the role $io^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'}$ (and similarly for other categories of basic paths that start and end in roots). Such roles can be seen as "aggregated paths". The "aggregated paths" will be unfolded afterwards into real paths, and the satisfaction of automata constraints by them will be verified by means of the *guided automata* (introduced in Definition 5). However, there exist cases of basic paths where the last element is unnamed, more precisely $(\mathtt{a},\mathtt{o})$-inner and $\mathtt{a}$-inner paths. In their cases we treat $\mathtt{a}^{\mathcal{I}}$ as the "virtual end" of ρ (and hence we link $\mathtt{a}^{\mathcal{I}}$ and $\mathtt{o}^{\mathcal{I}}$). To ensure that such an "aggregated path" can no longer be extended, we decorate the corresponding role with the dead-end symbol $\skull$. When constructing the guided automaton, the roles carrying $\skull$ will lead to states with no outgoing transitions (dead ends).

Definition 4. *Given a PSA* $\mathcal{A}$ *with the state-set* $\mathtt{Q}$*, the set of* ***$(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$-roles*** $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *is composed of the following role names (for all states* $\mathtt{q},\mathtt{q}' \in \mathtt{Q}$ *and names* $\mathtt{o}, \ddot{\mathtt{o}} \in \mathbf{N}_{\mathbf{I}}^{\mathcal{T}}$*):*

$$d^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'},\ i^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'_{\skull}},\ rt^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'},\ io^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'},\ oi^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'},\ i^{\mathcal{A},\mathtt{o}}_{\mathtt{q},\mathtt{q}'_{\skull}},\ by^{\mathcal{A},\mathtt{o},\ddot{\mathtt{o}}}_{\mathtt{q},\mathtt{q}'}.$$

An $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*-quasi-forest* $\mathcal{I}$ ***properly interprets*** $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *if its roles consist of pairs* p *of roots of* $\mathcal{I}$ *as indicated below.*

Role name r	Pair p	Condition
$d^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$	$(\mathrm{a}^{\mathcal{I}}, \mathrm{b}^{\mathcal{I}})$	$\mathrm{a}^{\mathcal{I}}\mathrm{b}^{\mathcal{I}} \models \mathcal{A}_{\mathtt{q},\mathtt{q}'}$
$i^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'_{\skull}}$	$(\mathrm{a}^{\mathcal{I}}, \mathrm{a}^{\mathcal{I}})$	$\mathrm{a}^{\mathcal{I}}$ is in $(\mathrm{I}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'})^{\mathcal{I}}$
$rt^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$	$(\mathrm{a}^{\mathcal{I}}, \mathrm{a}^{\mathcal{I}})$	$\mathrm{a}^{\mathcal{I}}$ is in $(\mathrm{RT}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'})^{\mathcal{I}}$
$io^{\mathcal{A},\circ}_{\mathtt{q},\mathtt{q}'}$	$(\mathrm{a}^{\mathcal{I}}, \circ^{\mathcal{I}})$	$\mathrm{a}^{\mathcal{I}}$ is in $(\mathrm{IO}^{\mathcal{A},\circ}_{\mathtt{q},\mathtt{q}'})^{\mathcal{I}}$
$oi^{\mathcal{A},\circ}_{\mathtt{q},\mathtt{q}'}$	$(\circ^{\mathcal{I}}, \mathrm{a}^{\mathcal{I}})$	$\mathrm{a}^{\mathcal{I}}$ is in $(\mathrm{OI}^{\mathcal{A},\circ}_{\mathtt{q},\mathtt{q}'})^{\mathcal{I}}$
$i^{\mathcal{A},\circ}_{\mathtt{q},\mathtt{q}'_{\skull}}$	$(\circ^{\mathcal{I}}, \mathrm{a}^{\mathcal{I}})$	$\mathrm{a}^{\mathcal{I}}$ is in $(\mathrm{I}^{\mathcal{A},\circ}_{\mathtt{q},\mathtt{q}'})^{\mathcal{I}}$
$by^{\mathcal{A},\circ,\ddot{\circ}}_{\mathtt{q},\mathtt{q}'}$	$(\circ^{\mathcal{I}}, \ddot{\circ}^{\mathcal{I}})$	$(\mathrm{By}^{\mathcal{A},\circ,\ddot{\circ}}_{\mathtt{q},\mathtt{q}'})^{\mathcal{I}}$ is non-empty

The size of $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *is polynomial w.r.t.* $|\mathcal{A}|\cdot|\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}|$. ◀

We stress that we interpreted role names from $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ based on the concepts from $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ rather than on the existence of certain paths realising $\mathcal{A}$. This allows us to verify their proper interpretation, independently from the verification of the proper interpretation of $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$-concepts. The roles from $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ can be seen as "shortcuts" aggregating fragments of runs of $\mathcal{A}$. To retrieve the runs, we invoke the $\mathcal{A}$-guided automaton that operates solely on such "shortcuts".

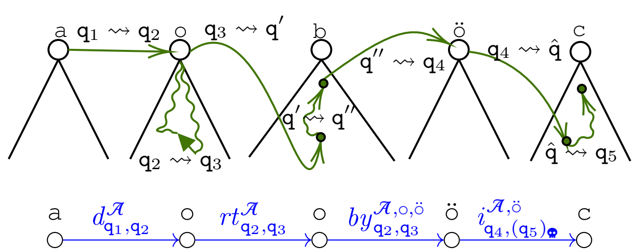
Definition 5. *For an NFA* $\mathcal{A}$ *with the state-set* $\mathtt{Q}$*, we define the* ***$\mathcal{A}$-guided NFA*** $\mathcal{B}$*. The set of states* $\mathtt{Q}'$ *of* $\mathcal{B}$ *consists of all* $\mathtt{q} \in \mathtt{Q}$ *and their fresh copies* $\mathtt{q}_{\skull}$*. The transitions in* $\mathcal{B}$ *have the form* $(\mathtt{q}, r, \mathtt{q}')$ *for all* $\mathtt{q}, \mathtt{q}' \in \mathtt{Q}'$*, and all* r *from* $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *labelled with the ordered pair* $(\mathtt{q}, \mathtt{q}')$*. By design, all states decorated with* ☠ *have no outgoing transitions.* ◀

The guided automaton $\mathcal{B}$ is of size polynomial in $|\mathcal{A}|\cdot|\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}|$, and it traverses only the clearings of quasi-forests. The following Lemma 7 reveals the desired property of $\mathcal{B}$.

Lemma 7. *Let* $\mathcal{A}$ *be an NFA with an* $\mathcal{A}$*-guided NFA* $\mathcal{B}$*, and* $\mathcal{I}$ *be an* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*-quasi-forest that properly interprets* $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ *and* $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$*. For all states* $\mathtt{q}, \mathtt{q}'$ *of* $\mathcal{A}$*:* $(\mathrm{Root} \sqcap \exists \mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top)^{\mathcal{I}} = (\mathrm{Root} \sqcap (\exists \mathcal{B}_{\mathtt{q},\mathtt{q}'}.\top \sqcup \exists \mathcal{B}_{\mathtt{q},\mathtt{q}'_{\skull}}.\top))^{\mathcal{I}}$. ◀

Its proof relies on either (a) shortening a path $\rho \models \mathcal{A}_{\mathtt{q},\mathtt{q}'}$ to its subsequence composed of all named elements, or (b) replacing any two consecutive elements in $\rho \models \mathcal{B}_{\mathtt{q},\mathtt{q}'}$ with the corresponding paths guaranteed by the roles from Def. 4.

Example 2. *Let* $\rho \coloneqq a^{\mathcal{I}}\cdot\circ^{\mathcal{I}}\cdot\rho_1\cdot\circ^{\mathcal{I}}\rho_2\cdot\ddot{\circ}^{\mathcal{I}}\cdot\rho_3 \models \mathcal{A}_{\mathtt{q}_1,\mathtt{q}_5}$ *be a path in a quasi-forest* $\mathcal{I}$ *(depicted below), and* $\mathcal{A}$ *be an NFA with states indicated by decorated* $\mathtt{q}$*. Suppose (i)* $a^{\mathcal{I}}\cdot\circ^{\mathcal{I}} \models \mathcal{A}_{\mathtt{q}_1,\mathtt{q}_2}$ *is* $(a, \circ)$*-direct, (ii)* $\circ^{\mathcal{I}}\cdot\rho_1\cdot\circ^{\mathcal{I}} \models \mathcal{A}_{\mathtt{q}_2,\mathtt{q}_3}$ *is an* $\circ$*-roundtrip, (iii)* $\circ^{\mathcal{I}}\rho_2\cdot\ddot{\circ}^{\mathcal{I}} \models \mathcal{A}_{\mathtt{q}_3,\mathtt{q}_4}$ *is an* $(b, \circ, \ddot{\circ})$*-bypass, and (iv)* $\ddot{\circ}^{\mathcal{I}}\cdot\rho_3 \models \mathcal{A}_{\mathtt{q}_4,\mathtt{q}_5}$ *is* $(c, \ddot{\circ})$*-inner. So,* $a^{\mathcal{I}} \in (\exists \mathcal{A}_{\mathtt{q}_1,\mathtt{q}_5}.\top)^{\mathcal{I}}$*.*



Consider now the $\mathcal{A}$*-guided* $\mathcal{B}$*, and observe that* $a^{\mathcal{I}}$ *is in* $(\exists \mathcal{B}_{\mathtt{q}_1,(\mathtt{q}_5)_{\skull}}.\top)^{\mathcal{I}}$*, as witnessed by the path* $a^{\mathcal{I}}\cdot\circ^{\mathcal{I}}\cdot\circ^{\mathcal{I}}\cdot\ddot{\circ}^{\mathcal{I}}\cdot c^{\mathcal{I}}$ *and the word* $d^{\mathcal{A}}_{\mathtt{q}_1,\mathtt{q}_2}\cdot rt^{\mathcal{A}}_{\mathtt{q}_2,\mathtt{q}_3}\cdot by^{\mathcal{A},\circ,\ddot{\circ}}_{\mathtt{q}_2,\mathtt{q}_3}\cdot i^{\mathcal{A},\ddot{\circ}}_{\mathtt{q}_4,(\mathtt{q}_5)_{\skull}}$. ◀

With the proviso that a quasi-forest $\mathcal{I}$ properly interprets both $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$ and $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$, Lemma 7 guarantees the desired interpretation of reachability concepts $\mathrm{Reach}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'}$.

Definition 6. *Let* $\mathcal{I}, \mathcal{A}, \mathcal{B}$ *be as in Lemma 7, and* $\mathtt{Q}$ *be the state-set of* $\mathcal{A}$*. The set of* ***$\mathcal{A}$-reachability-concepts*** $\mathbf{C}_{\rightsquigarrow}(\mathcal{A})$ *is* $\{\mathrm{Reach}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'} \mid \mathtt{q}, \mathtt{q}' \in \mathtt{Q}\}$*.* $\mathcal{I}$ ***properly interprets*** $\mathbf{C}_{\rightsquigarrow}(\mathcal{A})$ *if*

$$(\mathrm{Reach}^{\mathcal{A}}_{\mathtt{q},\mathtt{q}'})^{\mathcal{I}} = (\mathrm{Root} \sqcap (\exists \mathcal{B}_{\mathtt{q},\mathtt{q}'}.\top \sqcup \exists \mathcal{B}_{\mathtt{q},\mathtt{q}'_{\skull}}.\top))^{\mathcal{I}}$$

for all $\mathtt{q}, \mathtt{q}' \in \mathtt{Q}$*. We call* $\mathcal{I}$ ***virtually*** $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$***-decorated*** *if it properly interprets* $\mathbf{R}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$*, and* $\mathbf{C}_{\rightsquigarrow}(\mathcal{A})$*, and* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$***-decorated*** *if it additionally properly interprets* $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$. ◀

Relying on Lemma 5 and Lemma 7 we can show:

Lemma 8. *Let* $\mathcal{A}$ *be an NFA, and let* $\mathcal{I}$ *be an* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$*-decorated* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*-quasi-forest. For all elements* d *in* $\mathcal{I}$*, if* d *virtually satisfies* $\exists \mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top$ *then* d *satisfies* $\exists \mathcal{A}_{\mathtt{q},\mathtt{q}'}.\top$*.*

We conclude by showing an algorithmic result concerning (virtually) $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$-decorated quasi-forests. It exploits the fact that the regular path queries over databases can be evaluated in PTIME [Mendelzon and Wood, 1995, Lemma 3.1].

Lemma 9. *Let* $\mathcal{I}$ *be a finite interpretation with* $\mathrm{Root}^{\mathcal{I}} = \Delta^{\mathcal{I}}$*, and* $\mathcal{A}$ *be an NFA. We can then verify in time polynomial w.r.t.* $(|\mathcal{A}|\cdot|\mathcal{I}|)$ *whether* $\mathcal{I}$ *is virtually* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$*-decorated.* ◀

Lemma 9 tells us that if $\mathcal{I}$ is the clearing of some $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$-quasi-forest $\mathcal{J}$ that properly interprets $\mathbf{C}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$-concepts, then we can verify if $\mathcal{J}$ is $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \mathcal{A})$-decorated in PTIME.

5 Counting Decorations

We want to "relativise" number restrictions in the presence of nominals, so that in the suitably decorated quasi-forest models, the satisfaction of concepts of the form $(\geq n\ r).\top$ by the clearing can be decided solely based on the decoration of the clearing. Observe that for a given a root d of a quasi-forest $\mathcal{I}$ and a role name r, the set of r-successors of d can be divided into three groups: (a) the clearing, (b) the children of d, and (c) the descendants of roots (but only in case d is a nominal root). To relativise counting, we decorate each element of the clearing with the total number of their r-successors in categories (a) and (b), as well as the information, for each nominal $\circ$, on (c) how many of their descendants are r-successors of $\circ$.

Definition 7. *Fix* $r \in \mathbf{N_R}$*,* $n \in \mathbb{N}$*, and a finite* $\mathbf{N}_{\mathbf{I}}^{\mathcal{T}} \subseteq \mathbf{N_I}$*. The set of* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \geq n\ r)$***-counting-concepts*** $\mathbf{C}_{\#}(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \geq n\ r)$ *consists of concept names* $\mathrm{Clrng}^{r}_{\mathfrak{t}}$*,* $\mathrm{Chld}^{r}_{\mathfrak{t}}$*,* $\mathrm{Des}^{r,\circ}_{\mathfrak{t}}$ *for* $\circ \in \mathbf{N}_{\mathbf{I}}^{\mathcal{T}}$ *and thresholds* $\mathfrak{t}$ *of the form "*$=m$*" for* $0 \leq m \leq n$ *or "*$\geq n{+}1$*". All integers appearing in thresholds are encoded in binary. An* $(\mathbf{N}_{\mathbf{I}}^{\mathcal{A}}, \mathbf{N}_{\mathbf{I}}^{\mathcal{T}})$*-quasi-forest* $\mathcal{I}$ *is* $(\geq n\ r)$***-semi-decorated*** *if all roots* d *have unique thresholds* $\mathfrak{t}^{\mathrm{d}}_{cl}$*,* $\mathfrak{t}^{\mathrm{d}}_{ch}$*, and* $\mathfrak{t}^{\mathrm{d}}_{\circ}$ *for all* $\circ \in \mathbf{N}_{\mathbf{I}}^{\mathcal{T}}$*, for which* d *satisfies the concept*

$$\mathrm{Clrng}^{r}_{\mathfrak{t}^{\mathrm{d}}_{cl}} \sqcap \mathrm{Chld}^{r}_{\mathfrak{t}^{\mathrm{d}}_{ch}} \sqcap \textstyle\bigsqcap_{\circ \in \mathbf{N}_{\mathbf{I}}^{\mathcal{T}}} \mathrm{Des}^{r,\circ}_{\mathfrak{t}^{\mathrm{d}}_{\circ}}.$$

The above concepts are called the $(\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}, \geq n\ r)$***-descriptions of*** d*. Note that their size is polynomial w.r.t.* $|\mathbf{N}_{\mathbf{I}}^{\mathcal{T}}|\cdot\log_2(n)$.◀

Similarly to the previous section, we next introduce the notion of proper satisfaction. Intuitions regarding Definition 7 are provided in Example 3. We encourage the reader to read it.

Definition 8. *An* $(\mathbf{N_I^\mathcal{A}}, \mathbf{N_I^\mathcal{T}})$*-quasi-forest* $\mathcal{I}$ ***properly interprets*** $\mathbf{C}_\#(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$ *if* $\mathcal{I}$ *is* $({\geq}n\ r)$*-semi-decorated and for all roots* $a^\mathcal{I}$ *of* $\mathcal{I}$ *and* $o \in \mathbf{N_I^\mathcal{T}}$*, the root* $a^\mathcal{I}$ *belongs to:*

- $(\mathrm{Clrng}^r_\mathfrak{t})^\mathcal{I}$ *iff* $|\{\mathrm{d} \colon \mathrm{d} \in \mathrm{Root}^\mathcal{I}, (a^\mathcal{I}, \mathrm{d}) \in r^\mathcal{I}\}|\ \mathfrak{t}$,
- $(\mathrm{Chld}^r_\mathfrak{t})^\mathcal{I}$ *iff* $|\{\mathrm{d} \colon (a^\mathcal{I}, \mathrm{d}) \in r^\mathcal{I}, (a^\mathcal{I}, \mathrm{d}) \in \mathit{child}^\mathcal{I}\}|\ \mathfrak{t}$,
- $(\mathrm{Des}^{r,o}_\mathfrak{t})^\mathcal{I}$ *iff* $|\{\mathrm{d} \colon (o^\mathcal{I}, \mathrm{d}) \in r^\mathcal{I}, (a^\mathcal{I}, \mathrm{d}) \in (\mathit{child}^+)^\mathcal{I}\}|\ \mathfrak{t}$.

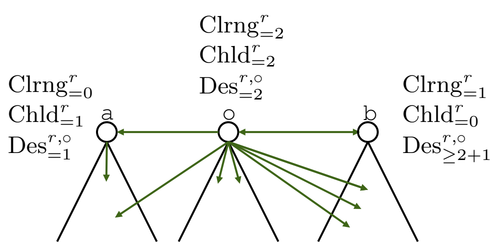
In this case we also say that $\mathcal{I}$ *is* $({\geq}n\ r)$***-decorated***. ◄

Based on the above definitions, we can easily express the intended behaviour of $\mathbf{C}_\#(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$-concepts in $\mathcal{ZOIQ}$, using $(=n\ r).\mathrm{C}$ as an abbreviation of $({\geq}n\ r).\mathrm{C} \sqcap \neg({\geq}n{+}1\ r).\mathrm{C}$. In the most difficult case, we describe $(\mathrm{Des}^{r,o}_{=42})^\mathcal{I}$ with:

$$\{\text{👻}\} \sqcap \exists \mathit{edge}^*. \big(\{o\} \sqcap (=42\ r).[\exists(\mathit{child}^-)^+.\{\text{👻}\}]\big)\,.$$

Lemma 10. *For every* $(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$*-description* C *there is a* $\mathcal{ZOIQ}$*-concept* $\mathrm{desc}(\mathrm{C})$ *(of size polynomial w.r.t.* $|\mathrm{C}|$*) that uses only individual names from* $\mathbf{N_I^\mathcal{T}} \cup \{\text{👻}\}$ *for a fresh* 👻*, s.t. for all* $(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$*-semi-decorated* $(\mathbf{N_I^\mathcal{A}}, \mathbf{N_I^\mathcal{T}})$*-quasi-forests* $\mathcal{I}$*:* $\mathcal{I}$ *is* $({\geq}n\ r)$*-decorated iff for every root* $a^\mathcal{I}$ *of* $\mathcal{I}$ *and its* $(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$*-description* C *we have* $a^\mathcal{I} \in (\mathrm{desc}(\mathrm{C})[\text{👻}/a])^\mathcal{I}$.◄

Example 3. *Let us consider an* $(\{a, b\}, \{o\})$*-quasi-forest* $\mathcal{I}$ *sketched below and a role name* r *depicted as a green arrow. As suggested by the drawing, (i)* $a^\mathcal{I}$ *has no* r*-successors among the clearing of* $\mathcal{I}$ *and precisely one* r*-successor among its children, (ii)* $o^\mathcal{I}$ *has two* r*-successors among the clearing of* $\mathcal{I}$*, it has precisely two* r*-successor among its children, one* r*-successor that is a descendant of* $a^\mathcal{I}$*, and three* r*-successors that are descendants of* $b^\mathcal{I}$*, and (iii)* $b^\mathcal{I}$ *has precisely one* r*-successor inside the clearing of* $\mathcal{I}$ *and no other* r*-successors.*



Suppose now that $\mathcal{I}$ *is* $(\{o\}, {\geq}2\ r)$*-decorated. This implies:*

- $a^\mathcal{I} \in (\mathrm{Clrng}^r_{=0} \sqcap \mathrm{Chld}^r_{=1} \sqcap \mathrm{Des}^{r,o}_{=1})^\mathcal{I}$.
- $o^\mathcal{I} \in (\mathrm{Clrng}^r_{=2} \sqcap \mathrm{Chld}^r_{=2} \sqcap \mathrm{Des}^{r,o}_{=2})^\mathcal{I}$.
- $b^\mathcal{I} \in (\mathrm{Clrng}^r_{=1} \sqcap \mathrm{Chld}^r_{=0} \sqcap \mathrm{Des}^{r,o}_{\geq 2+1})^\mathcal{I}$. ◄

Lemma 10 rephrases the notion of proper satisfaction in the language of $\mathcal{ZOIQ}$-concepts. Call a quasi-forest $\mathcal{I}$ **virtually** $(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$**-decorated** if $\mathcal{I}$ is $(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$-semi-decorated and properly interprets concepts of the form $\mathrm{Clrng}^r_\mathfrak{t}$. We have:

Lemma 11. *For a finite interpretation* $\mathcal{I}$ *we can test in polynomial time w.r.t.* $|\mathcal{I}|$ *if* $\mathcal{I}$ *is virtually* $(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$*-decorated.* ◄

We use the numbers appearing in $(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$-descriptions labelling the roots of $({\geq}n\ r)$-decorated quasi-forests to decide the satisfaction of $({\geq}n\ r).\top$. We first describe it with an example. Suppose that we want to verify if a non-nominal root d from $\mathcal{I}$ satisfies $({\geq}3\ r)$ based on its labels $\mathrm{Clrng}^r_{=1}$ and $\mathrm{Chld}^r_{\geq 3+1}$. It suffices to check if $1 + (3+1)$ is at least 3. For the nominals roots o, we additionally take all the concepts $\mathrm{Des}^{r,o}_\mathfrak{t}$ into account, that are spread across the whole clearing.

An element d from a finite $\mathcal{I}$ **virtually satisfies** $({\geq}n\ r).\top$, whenever d satisfies $({\geq}n\ r).\top$ in every $({\geq}n\ r)$-decorated $(\mathbf{N_I^\mathcal{A}}, \mathbf{N_I^\mathcal{T}})$-quasi-forest with the clearing equal to $\mathcal{I}$. We show:

Lemma 12. *For a finite virtually* $(\mathbf{N_I^\mathcal{T}}, {\geq}n\ r)$*-decorated interpretation* $\mathcal{I}$ *with* $\Delta^\mathcal{I} = \mathrm{Root}^\mathcal{I}$ *and* $\mathrm{d} \in \mathrm{Root}^\mathcal{I}$ *we can test in time polynomial in* $|\mathcal{I}|$ *if* d *virtually satisfies* $({\geq}n\ r).\top$. ◄

6 Elegant Models and Their Summaries

In this section we benefit from various decorations introduced in the previous sections to design a succinct way of representing quasi-forest models of $\mathcal{ZOIQ}$-KBs, dubbed *summaries*. From now on we will focus only on KBs in Scott's normal form (*i.e.* with TBoxes in Scott's normal form) and on certain class of *elegant* models introduced in Definition 9.

Definition 9. *Let* $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ *be a* $\mathcal{ZOIQ}$*-KB in Scott's normal form and let* $\mathcal{I}$ *be its model.*
We call $\mathcal{I}$ ***elegant*** *if it satisfies all the conditions below:*

(i) $\mathcal{I}$ *is a canonical quasi-forest model of* $\mathcal{K}$*,*
(ii) $\mathcal{I}$ *is* $(\mathsf{ind}(\mathcal{T}), {\geq}n\ r)$*-decorated for all number restrictions* $({\geq}n\ r).\top$ *from* $\mathcal{T}$*,*
(iii) $\mathcal{I}$ *is* $(\mathsf{ind}(\mathcal{T}), \mathcal{A})$*-decorated for all automata* $\mathcal{A}$ *from* $\mathcal{T}$*,*
(iv) $\mathcal{I}$ *interprets all concept and role names that do not appear in* $\mathcal{K}$*, the set* $\{\mathrm{Root}, \mathit{edge}, \mathit{child}, \mathit{id}\}$*, and in mentioned decorations, as the empty set.* ◄

Invoking the normal form lemma [Appendix A], Lemma 1, as well as the definitions of proper interpretation, we obtain:

Lemma 13. *For every* $\mathcal{ZOIQ}$*-TBox* $\mathcal{T}$ *we can compute in* PTIME *a* $\mathcal{ZOIQ}$*-TBox* $\mathcal{T}'$ *in Scott's normal form that possibly such that for every ABox* $\mathcal{A}$ *we have: (i)* $(\mathcal{A}, \mathcal{T})$ *is quasi-forest satisfiable iff* $(\mathcal{A}, \mathcal{T}')$ *has an elegant model, and (ii) for every UCQ* q *using only concepts and roles present in* $\mathcal{T}$ *we have that* $(\mathcal{A}, \mathcal{T})$ *has a canonical quasi-forest model violating* q *iff* $(\mathcal{A}, \mathcal{T}')$ *has an elegant model violating* q. ◄

To-be-defined summaries are nothing more than ABoxes representing the full descriptions of clearings of elegant models.

Definition 10. *Let* $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ *be a* $\mathcal{ZOIQ}$*-KB in Scott's normal form. A* $\mathcal{K}$***-summary*** $\mathcal{S}$ *is any* $\subseteq$*-minimal ABox satisfying, for all names* $a, b \in \mathsf{ind}(\mathcal{K})$*, all the conditions below.*

(I) $\mathcal{S}$ *contains either* $a \approx b$ *or* $\neg(a \approx b)$*.*
(II) For all concept names A *appearing in* $\mathcal{K}$ *we have that* $\mathcal{S}$ *contains either* $\mathrm{A}(a)$ *or* $\neg\mathrm{A}(a)$*.*
(III) For all NFA $\mathcal{A}$ *from* $\mathcal{K}$*, and all concept names* A *from* $\mathbf{C}(\mathsf{ind}(\mathcal{T}), \mathcal{A}) \cup \mathbf{C}_{\rightsquigarrow}(\mathcal{A})$*,* $\mathcal{S}$ *contains* $\mathrm{A}(a)$ *or* $\neg\mathrm{A}(a)$*.*
(IV) For all role names r *appearing in* $\mathcal{K}$ *we have that* $\mathcal{S}$ *contains either* $r(a, b)$ *or* $\neg r(a, b)$*.*
(V) For all NFA $\mathcal{A}$ *from* $\mathcal{K}$*, and all role names* r *from* $\mathbf{R}(\mathsf{ind}(\mathcal{T}), \mathcal{A})$*, the ABox* $\mathcal{S}$ *contains* $r(a, b)$ *or* $\neg r(a, b)$*.*
(VI) For all number restrictions $({\geq}n\ r).\top$ *from* $\mathcal{K}$*, and for all* $o \in \mathsf{ind}(\mathcal{T})$*, there are thresholds* $\mathfrak{t}^a_{cl}$*,* $\mathfrak{t}^a_{ch}$*, and* $\mathfrak{t}^a_o$ *in* $\{=m, {\geq}n{+}1 \mid 0 \leq m \leq n\}$ *for which* $\mathcal{S}$ *contains* $\mathrm{Clrng}^r_{\mathfrak{t}^a_{cl}}(a)$*,* $\mathrm{Chld}^r_{\mathfrak{t}^a_{ch}}(a)$*, and* $\mathrm{Des}^{r,o}_{\mathfrak{t}^a_o}(a)$*.*
(VII) $\mathrm{Root}(a) \in \mathcal{S}$ *and* $\mathit{edge}(a, b) \in \mathcal{S}$*.*

In the case when only a $\mathcal{ZOIQ}$*-TBox* $\mathcal{T}$ *is given, we define* $\mathcal{T}$***-ghost-summaries*** *as* $(\{\text{👻} \approx \text{👻}\}, \mathcal{T})$*-summaries.* ◄

Invoking the previously-established bounds on the number and sizes of concepts and roles occurring in decorations (*i.e.* the ones stated at the end of Definitions 3, 4, and 7), we infer:

Lemma 14. *Let $\mathcal{K}$ and $\mathcal{T}$ be, respectively, a $\mathcal{ZOIQ}$-KB and a $\mathcal{ZOIQ}$-TBox, both in Scott's normal form. We have that every $\mathcal{K}$-summary is of size polynomial in $|\mathcal{K}|$. Moreover, there are exponentially (in $|\mathcal{T}|$) many $\mathcal{T}$-ghost-summaries and each of them is of size polynomial w.r.t. $|\mathcal{T}|$.* ◀

While a $\mathcal{K}$-summary can be easily extracted from the clearing of any elegant model of $\mathcal{K}$, the converse direction requires two extra assumptions, dubbed *clearing-* and *subtree-consistency*.

7 Consistent Summaries

The notion of *clearing-consistency* ensures that a given summary $\mathcal{S}$ is a good candidate for the clearing of some elegant model of $\mathcal{K}$, namely (i) $\mathcal{S}$ does not violate "local" constraints from $\mathcal{K}$, (ii) $\mathcal{S}$ is virtually decorated for decorations involving number restrictions and automata, and that (iii) every element from $\mathcal{S}$ virtually satisfies all concepts involving automata or number restrictions from $\mathcal{K}$. Below, $\mathcal{I}_{\mathcal{S}}$ denotes the minimal interpretation that corresponds to and satisfies the ABox $\mathcal{S}$.

Definition 11. *Let $\mathcal{K} \coloneqq (\mathcal{A}, \mathcal{T})$ be a $\mathcal{ZOIQ}$-KB in Scott's normal form, and let $\mathcal{S}$ be a $\mathcal{K}$-summary. Let $\mathcal{T}_{loc}$ be the TBox composed of all GCIs from $\mathcal{T}$ except for the ones concerning NFAs and number restrictions. We say that $\mathcal{S}$ is* ***clearing-consistent*** *if $\mathcal{I}_{\mathcal{S}}$ satisfies $(\mathcal{A}, \mathcal{T}_{loc})$, and:*
• For all GCIs $\mathrm{A} \equiv \exists\mathcal{A}_{\mathsf{q},\mathsf{q}'}.\top$ from $\mathcal{T}$: (i) $\mathcal{I}_{\mathcal{S}}$ is virtually $(\mathsf{ind}(\mathcal{T}), \mathcal{A})$-decorated, and (ii) for all $a \in \mathsf{ind}(\mathcal{K})$, $\mathrm{A}(a) \in \mathcal{S}$ iff a virtually satisfies $\exists\mathcal{A}_{\mathsf{q},\mathsf{q}'}.\top$ in $\mathcal{I}_{\mathcal{S}}$.
• For all GCIs $\mathrm{A} \equiv ({\geq}n\ r).\top$ from $\mathcal{T}$: (i) $\mathcal{I}_{\mathcal{S}}$ is virtually $(\mathsf{ind}(\mathcal{T}), {\geq}n\ r)$-decorated, and (ii) for all $a \in \mathsf{ind}(\mathcal{K})$, we have that $\mathrm{A}(a) \in \mathcal{S}$ iff a virtually satisfies $({\geq}n\ r).\top$ in $\mathcal{I}_{\mathcal{S}}$. ◀

Based on Lemma 9 and Lemma 12 we can conclude that deciding clearing consistency can be done in PTIME.

Lemma 15. *For $\mathcal{K}$ and $\mathcal{S}$ as in Def. 11 we can decide in time polynomial w.r.t. $|\mathcal{K}| \cdot |\mathcal{S}|$ whether $\mathcal{S}$ is clearing consistent.* ◀

The second required notion is the *subtree-consistency*. To decide whether a given summary $\mathcal{S}$ extends to a model of $\mathcal{K}$, we need to check, for all elements d of $\mathcal{S}$, the existence of a suitable forest satisfying (a relativised) $\mathcal{K}$ and fulfilling all the premises given by the decorations of d. This is achieved by crafting a suitable $\mathcal{ZOIQ}$-KB and testing whether it has a quasi-forest model of a suitably bounded branching. To make such a KB dependent only on the TBox, we are going to use the ghost variable 👻 in place of the intended element d.

Definition 12. *Let $\mathcal{T}$ be a $\mathcal{ZOIQ}$-TBox in Scott's normal form and $\mathcal{S}$ be a $\mathcal{T}$-ghost-summary. $\mathcal{S}$ is* ***$\mathcal{T}$-subtree-consistent*** *if the $\mathcal{ZOIQ}$-KB $\mathcal{K}_{\mathcal{S},👻} \coloneqq (\mathcal{S}, \mathcal{T}_{loc} \cup \mathcal{T}_{aut} \cup \mathcal{T}_{cnt})$ is quasi-forest satisfiable, where $\mathcal{T}_{loc}$ is as in Definition 11 and:*
• $\mathcal{T}_{aut}$ consists of $\{👻\} \sqsubseteq \mathrm{comdsc}(\mathsf{ind}(\mathcal{T}), \mathcal{A})$ and $\mathrm{A} \equiv \mathrm{rel}(\mathsf{ind}(\mathcal{T}), \mathcal{A}_{\mathsf{q},\mathsf{q}'})$ for all GCIs $\mathrm{A} \equiv \exists\mathcal{A}_{\mathsf{q},\mathsf{q}'}.\top$ from $\mathcal{T}$.
• $\mathcal{T}_{cnt}$ consists of the GCIs $\neg\mathrm{Root} \sqcap \mathrm{A} \equiv \neg\mathrm{Root} \sqcap ({\geq}n\ r).\top$ as well as $\{👻\} \sqsubseteq \mathrm{desc}\left(\mathrm{Chld}^{r}_{\mathbf{t}^{👻}_{ch}} \sqcap \bigsqcap_{o \in \mathsf{ind}(\mathcal{T})} \mathrm{Des}^{r;o}_{\mathbf{t}^{👻}_{o}}\right)$, for the unique concepts $\mathrm{Chld}^{r}_{\mathbf{t}^{👻}_{ch}}$, $\mathrm{Des}^{r;o}_{\mathbf{t}^{👻}_{o}}$ satisfied by 👻 in $\mathcal{I}_{\mathcal{S}}$, for all GCIs $\mathrm{A} \equiv ({\geq}n\ r).\top$ appearing in $\mathcal{T}$. ◀

The crucial property concerning the notion of $\mathcal{T}$-subtree-consistency as well as the set $\mathbf{S}_{\mathcal{T}}$ of all $\mathcal{T}$-subtree-consistent $\mathcal{T}$-ghost-summaries is provided next. Its proof relies on the bounds on the concepts from Lemma 5, Lemma 6, and Definition 7, as well as the exponential time algorithm for deciding quasi-forest-satisfiability of $\mathcal{ZOIQ}$-KBs from Lemma 2.

Lemma 16. *For a $\mathcal{ZOIQ}$-TBox $\mathcal{T}$ in Scott's normal form, one can decide in time exponential w.r.t. $|\mathcal{T}|$ if a given $\mathcal{T}$-ghost-summary is $\mathcal{T}$-subtree-consistent. Moreover, the set $\mathbf{S}_{\mathcal{T}}$ of all $\mathcal{T}$-subtree-consistent $\mathcal{T}$-ghost-summaries has size exponential in $|\mathcal{T}|$ and is computable in time exponential in $|\mathcal{T}|$.* ◀

We now lift the definition of subtree-consistency from ghost summaries to arbitrary $\mathcal{K}$-summaries by producing a ghost summary per each individual name $\mathtt{a}$ mentioned in $\mathcal{K}$. The idea is simple: we first restrict $\mathcal{S}$ to nominals and the selected name $\mathtt{a}$, and then replace $\mathtt{a}$ with the ghost variable 👻 (we must be a bit more careful if $\mathtt{a}$ is itself a nominal). This produces a ghost summary, for which the notion of subtree-consistency is already well-defined. The main benefit of testing subtree-consistency of such a summary is that this guarantees the existence of a quasi-forest containing a subtree rooted at 👻 that we can afterwards "plug in" to $\mathtt{a}$ in $\mathcal{S}$ in order to produce a full model of $\mathcal{K}$ from $\mathcal{S}$. This is formalised next.

Definition 13. *Let $\mathcal{T}$ and $\mathbf{S}_{\mathcal{T}}$ be as in Lemma 16. For a $\mathcal{ZOIQ}$-KB $\mathcal{K} \coloneqq (\mathcal{A}, \mathcal{T})$ and a $\mathcal{K}$-summary $\mathcal{S}$ we say that $\mathcal{S}$ is* ***consistent*** *if $\mathcal{S}$ is clearing-consistent and for every name $a \in \mathsf{ind}(\mathcal{K})$ the* ***$(\mathcal{T}, \mathcal{S}, \mathtt{a})$-ghost-summary $\mathcal{S}_{\mathtt{a}}$*** *belongs to $\mathbf{S}_{\mathcal{T}}$. $\mathcal{S}_{a}$ is defined as the sum of (i) $\mathcal{S}$ restricted to $\mathsf{ind}(\mathcal{T})$, (ii) $\mathcal{S}$ restricted to $\mathsf{ind}(\mathcal{T}) \cup \{a\}$ with all occurrences of a replaced with 👻, and (iii) $\{a \approx 👻, 👻 \approx a\}$ in case $a \in \mathsf{ind}(\mathcal{T})$.* ◀

It can be readily verified that all $\mathcal{K}$-summaries constructed from elegant models of $\mathcal{ZOIQ}$-KBs $\mathcal{K} \coloneqq (\mathcal{A}, \mathcal{T})$ are consistent. For the opposite direction, we define a suitable notion of a merge in order to "combine" a clearing-consistent summary $\mathcal{S}$ with relevant subtrees provided by the subtree-consistency of $(\mathcal{T}, \mathcal{S}, \mathtt{a})$-ghost-summaries for all $\mathtt{a} \in \mathsf{ind}(\mathcal{K})$. The interpretation constructed in this way becomes an elegant model of $\mathcal{K}$.

Lemma 17. *A $\mathcal{ZOIQ}$-KB in Scott's normal form is quasi-forest satisfiable iff there exists a consistent $\mathcal{K}$-summary.* ◀

8 Deciding Existence of Quasi-Forest Models

Based on Lemma 17 and all the other lemmas presented in the previous section, we can finally design an algorithm for deciding whether an input $\mathcal{ZOIQ}$-KB is quasi-forest satisfiable.

Algorithm 1: Quasi-Forest Satisfiability in $\mathcal{ZOIQ}$

```
Input: A ZOIQ-KB K := (A, T).
Output: True iff K is quasi-forest satisfiable.
Turn T into Scott's normal form.                 // L. 13.
Compute the set S_T.                             // L. 16.
Guess a K-summary S.                             // L. 14.
Return False if S is clearing-consistent.        // L. 15.
Foreach a ∈ ind(K) do compute the (T, S, a)-ghost-
  -summary S_a and return False if S_a ∉ S_T.
Return True.
```

Lemma 18. *Algorithm 1 returns* `True` *if and only if the input $\mathcal{ZOIQ}$-KB $\mathcal{K}$ has an elegant model. Moreover, there exist polynomial function* p *and an exponential function* e *for which Algorithm 1 can be implemented with a nondeterministic Turing machine of running time bounded, for every input $\mathcal{K} := (\mathcal{A}, \mathcal{T})$, by $\mathrm{e}(|\mathcal{T}|) + \mathrm{p}(|\mathcal{K}|) + \mathrm{p}(|\mathcal{K}|) \cdot \mathrm{e}(|\mathcal{T}|)$.* ◀

Suppose now that a TBox $\mathcal{T}$ is fixed and only an ABox $\mathcal{A}$ is given as the input. Then the first two steps of Algorithm 1 are independent from the input. The same holds for the verification of whether a given ghost summary belongs to the pre-computed set $\mathbf{S}_{\mathcal{T}}$. The desired algorithm is given below.

Algorithm 2: Deciding Quasi-Forest Satisfiability in $\mathcal{ZOIQ}$ w.r.t. Data Complexity

---

Input: An ABox $\mathcal{A}$.
Parameters: $\mathcal{ZOIQ}$-TBox $\mathcal{T}$ already in Scott's normal form and a pre-computed set $\mathbf{S}_{\mathcal{T}}$.
Output: `True` iff $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ is quasi-forest sat.

Guess a $\mathcal{K}$-summary $\mathcal{S}$. `// NP in` $|\mathcal{K}|$`, L. 14.`
Return `False` if $\mathcal{S}$ is clearing-consistent.
`// PTime w.r.t.` $|\mathcal{K}|\cdot|\mathcal{S}|$ `by L. 15.`
Foreach $\mathsf{a} \in \mathsf{ind}(\mathcal{K})$ **do** compute the $(\mathcal{T}, \mathcal{S}, \mathsf{a})$-ghost-summary $\mathcal{S}_{\mathsf{a}}$ and return `False` if $\mathcal{S}_{\mathsf{a}} \notin \mathbf{S}_{\mathcal{T}}$.
`// PTime w.r.t.` $|\mathcal{K}| \cdot (|\mathcal{K}|+|\mathbf{S}_{\mathcal{T}}|)$
Return `True`.

We employ Lemma 18, and based on Algorithm 2 we get:

Theorem 1. *For every $\mathcal{ZOIQ}$-TBox $\mathcal{T}$, there is an* NP *procedure (parametrised by $\mathcal{T}$) that for an input ABox $\mathcal{A}$ decides if the $\mathcal{ZOIQ}$-KB $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ is quasi-forest satisfiable.* ◀

Recall that the maximal decidable fragments of $\mathcal{ZOIQ}$ have the elegant model property (Lemma 13). We conclude:

Theorem 2. *The satisfiability problem for $\mathcal{ZIQ}$, $\mathcal{ZOQ}$, and $\mathcal{ZOI}$ is* NP*-complete w.r.t. the data complexity.* ◀

For expressive logics from the $\mathcal{SR}$ family, the logical core of OWL2, it is known [Calvanese *et al.*, 2009, Prop. 5.1] that any TBox in $\mathcal{SRIQ}$, $\mathcal{SROQ}$ or $\mathcal{SROI}$ can be rewritten into $\mathcal{ZIQ}$, $\mathcal{ZOQ}$ or $\mathcal{ZOI}$. Hence, we reprove the following result:

Corollary 1. *The satisfiability problem for $\mathcal{SRIQ}$, $\mathcal{SROQ}$, and $\mathcal{SROI}$ is* NP*-complete w.r.t. the data complexity.* ◀

9 Application: Entailment of Rooted Queries

We adapt Algorithm 1 to derive coNExpTime-completeness of the entailment problem of (unions of) rooted conjunctive queries over $\mathcal{ZIQ}$-KBs, generalising previous results on $\mathcal{SHIQ}$ [Lutz, 2008b, Thm. 2]. We focus on the dual problem: "Given a (union of) rooted CQs q and $\mathcal{ZIQ}$-KB $\mathcal{K} := (\mathcal{A}, \mathcal{T})$, is there a model of $\mathcal{K}$ that violates q?" and show its NExpTime-completeness. As we work with $\mathcal{ZIQ}$, by Lemma 13 we can assume that the input KB $\mathcal{K}$ is in Scott's normal form, and the intended model $\mathcal{I}$ violating q is elegant (in particular, is N-bounded for an N exponential in $|\mathcal{T}|$). The crucial observation is that whenever q matches $\mathcal{I}$, all the elements from the image of a match are at the depth at most $|q|$ (as q has at least one individual name and is connected). Hence, it suffices to construct an "initial segment" of $\mathcal{I}$ of depth at most $|q|$ and degree at most N, and check if (i) q does not match it, and (ii) it can be extended to the full model of $\mathcal{K}$.

Definition 14. *Let $\mathrm{R}, \mathrm{C}, \mathrm{D} \in \mathbb{N}$. An $(\mathrm{R}, \mathrm{C}, \mathrm{D})$**-forest** $\mathcal{F}$ is a prefix-closed set of non-empty words from $\mathbb{Z}_{\mathrm{R}} \cdot (\mathbb{Z}_{\mathrm{C}})^+$ of length at most D ($\mathbb{Z}_n$ denotes the set of integers modulo n). The number R indicates the total number of roots of $\mathcal{F}$, C denotes the maximal number of children per each element, and D indicates the maximal depth.*

*Treating $\mathcal{F}$ as a set of individual names, we define an $(\mathcal{F}, \mathcal{K})$**-initial segment** of a $\mathcal{ZOIQ}$-KB $\mathcal{K}$ as any summary $\mathcal{S}$ of $\mathcal{K} \cup \{a \not\approx b \mid a, b \in \mathcal{F}, a \neq b\}$ such that:*
- *For every $a \in \mathsf{ind}(\mathcal{K})$ there is $b \in \mathcal{F} \cap \mathbb{N}$ with $a \approx b$ in $\mathcal{S}$, and for every $b \in \mathcal{F} \cap \mathbb{N}$ there is $a \in \mathsf{ind}(\mathcal{K})$ with $a \approx b$ in $\mathcal{S}$.*
- *$((=0\ child).\top)\ (a)$ for all $a \in \mathcal{F}$ that are not leaves of $\mathcal{F}$.*
- *$\neg r(a, b)$ for all role names r appearing in $\mathcal{K}$ and all $a \neq b$ from $\mathcal{F}$ such that a is not a child of b in $\mathcal{F}$ (or vice-versa).* ◀

The above conditions are needed to represent a forest $\mathcal{F}$ inside the clearing of the intended models of $\mathcal{S}$. The first item of Definition 14 guarantees the proper behaviour of roots. The second item guarantees that the children of elements in the initial segment are precisely the ones that are explicitly mentioned there. Finally, the last item ensures that the "tree-likeness" of the initial segment is not violated.

Lemma 19. *Let $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ be a $\mathcal{ZIQ}$-KB in Scott's normal form and let $q := \bigvee_i q_i$ be a union of rooted CQs. We have that $\mathcal{K} \not\models q$ if and only if there is an $(\mathcal{F}, \mathcal{K})$-initial segment $\mathcal{S}$ such that (i) $\mathcal{F}$ is an $(\mathrm{R}, \mathrm{C}, \mathrm{D})$-forest for an R bounded by $|\mathsf{ind}(\mathcal{K})|$, C exponential in $|\mathcal{T}|$, and D bounded by $|q|$, (ii) $(\mathcal{S}, \mathcal{T})$ is quasi-forest satisfiable, and (iii) $\mathcal{I}_{\mathcal{S}} \not\models q$.* ◀

Note that the forest $\mathcal{F}$ described above is exponential w.r.t. $|\mathcal{K}|$, and thus the initial segment $\mathcal{S}$ can be "guessed" in NExp.

Algorithm 3: Rooted Query Entailment in $\mathcal{ZIQ}$

Input: A $\mathcal{ZIQ}$-KB $\mathcal{K} := (\mathcal{A}, \mathcal{T})$ and a union $q := \bigvee_{i=1}^{n} q_i$ of rooted conjunctive queries.
Output: `True` if and only if $\mathcal{K} \not\models q$.

Turn $\mathcal{T}$ into Scott's normal form.
Guess $(\mathcal{F}, \mathcal{K})$-initial segment $\mathcal{S}$ of $\mathcal{K}$ as in Lemma 19.
Foreach $1 \leq i \leq n$ return `False` if $\mathcal{I}_{\mathcal{S}} \models q_i$.
Use Algorithm 1 to verify that $(\mathcal{S}, \mathcal{T})$ is quasi-forest satisfiable and return `True` if this is indeed the case.

Lemma 20. *Algorithm 3 returns* `True` *if and only if the input $\mathcal{ZOIQ}$-KB $\mathcal{K}$ does not entail the input query q. Moreover, Algorithm 3 can can be implemented with a nondeterministic Turing machine of running time bounded by some function bounded exponentially w.r.t. $|q|$ and $|\mathcal{K}|$.* ◀

Based on Lemma 20 we can conclude:

Theorem 3. *The entailment problem for unions of rooted CQs over $\mathcal{ZIQ}$-KBs is co*NExpTime*-complete.* ◀

The coNExpTime-hardness of rooted entailment holds already for $\mathcal{ALCI}$ [Lutz, 2008a, Thm. 1]. Interestingly enough, we establish [Appendix L] a novel lower bound for $\mathcal{ALC}\mathsf{Self}$.

Theorem 4. *The rooted conjunctive query entailment problem is co*NExpTime*-hard for $\mathcal{ALC}\mathsf{Self}$ (thus also for $\mathcal{ZIQ}$).* ◀

Acknowledgements

This work was supported by the ERC Grant No. 771779 (DeciGUT). Full version of this paper is available on ArXiV.

References

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

[Bednarczyk and Kieroński, 2022] Bartosz Bednarczyk and Emanuel Kieroński. Finite Entailment of Local Queries in the Z Family of Description Logics. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 5487–5494. AAAI Press, 2022.

[Bednarczyk and Rudolph, 2019] Bartosz Bednarczyk and Sebastian Rudolph. Worst-Case Optimal Querying of Very Expressive Description Logics with Path Expressions and Succinct Counting. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 1530–1536. ijcai.org, 2019.

[Bienvenu *et al.*, 2014] Meghyn Bienvenu, Diego Calvanese, Magdalena Ortiz, and Mantas Simkus. Nested Regular Path Queries in Description Logics. In Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press, 2014.

[Calvanese *et al.*, 2009] Diego Calvanese, Thomas Eiter, and Magdalena Ortiz. Regular Path Queries in Expressive Description Logics with Nominals. In Craig Boutilier, editor, *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009*, pages 714–720, 2009.

[Calvanese *et al.*, 2016] Diego Calvanese, Magdalena Ortiz, and Mantas Simkus. Verification of Evolving Graph-structured Data under Expressive Path Constraints. In Wim Martens and Thomas Zeume, editors, *19th International Conference on Database Theory, ICDT 2016, Bordeaux, France, March 15-18, 2016*, volume 48 of *LIPIcs*, pages 15:1–15:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

[Hitzler *et al.*, 2012] Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph. *OWL 2 Web Ontology Language Primer (Second Edition)*. World Wide Web Consortium (W3C), December 2012.

[Jung *et al.*, 2018] Jean Christoph Jung, Carsten Lutz, Mauricio Martel, and Thomas Schneider. Querying the Unary Negation Fragment with Regular Path Expressions. In Benny Kimelfeld and Yael Amsterdamer, editors, *21st International Conference on Database Theory, ICDT 2018, March 26-29, 2018, Vienna, Austria*, volume 98 of *LIPIcs*, pages 15:1–15:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[Kazakov, 2008] Yevgeny Kazakov. RIQ and SROIQ are harder than SHOIQ. In Gerhard Brewka and Jérôme Lang, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*, pages 274–284. AAAI Press, 2008.

[Lutz, 2008a] Carsten Lutz. The Complexity of Conjunctive Query Answering in Expressive Description Logics. In Alessandro Armando, Peter Baumgartner, and Gilles Dowek, editors, *Automated Reasoning, 4th International Joint Conference, IJCAR 2008, Sydney, Australia, August 12-15, 2008, Proceedings*, volume 5195 of *Lecture Notes in Computer Science*, pages 179–193. Springer, 2008.

[Lutz, 2008b] Carsten Lutz. Two Upper Bounds for Conjunctive Query Answering in SHIQ. In Franz Baader, Carsten Lutz, and Boris Motik, editors, *Proceedings of the 21st International Workshop on Description Logics (DL2008), Dresden, Germany, May 13-16, 2008*, volume 353 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.

[Mendelzon and Wood, 1995] Alberto O. Mendelzon and Peter T. Wood. Finding Regular Simple Paths in Graph Databases. *SIAM Journal on Computing*, 24(6):1235–1258, 1995.

[Ortiz and Simkus, 2012] Magdalena Ortiz and Mantas Simkus. Reasoning and Query Answering in Description Logics. In Thomas Eiter and Thomas Krennwallner, editors, *Reasoning Web. Semantic Technologies for Advanced Query Answering - 8th International Summer School 2012, Vienna, Austria, September 3-8, 2012. Proceedings*, volume 7487 of *Lecture Notes in Computer Science*, pages 1–53. Springer, 2012.

[Ortiz, 2010] Magdalena Ortiz. *Query Answering in Expressive Description Logics: Techniques and Complexity Results*. PhD thesis, TU Wien, AT, 2010.

[Pratt-Hartmann, 2009] Ian Pratt-Hartmann. Data-Complexity of the Two-Variable Fragment with Counting Quantifiers. *Information and Computation*, 207(8):867–888, 2009.

[Sipser, 2013] Michael Sipser. *Introduction to the Theory of Computation*. Course Technology, Boston, MA, Third edition, 2013.

[Vardi, 1982] Moshe Y. Vardi. The Complexity of Relational Query Languages (Extended Abstract). In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing (STOC 1982)*, STOC '82, page 137–146, New York, NY, USA, 1982. Association for Computing Machinery.