TECHNISCHE UNIVERSITÄT DRESDEN

Technische Universität Dresden
Institute for Theoretical Computer Science
Chair for Automata Theory

LTCS–Report

Conjunctive Query Answering in Rough $\mathcal{EL}$

Rafael Peñaloza Veronika Thost Anni-Yasmin Turhan

LTCS-Report 14-04

Postal Address:
Lehrstuhl für Automatentheorie
Institut für Theoretische Informatik
TU Dresden
01062 Dresden

http://lat.inf.tu-dresden.de

Visiting Address:
Nöthnitzer Str. 46
Dresden

Conjunctive Query Answering in Rough $\mathcal{EL}$

Rafael Peñaloza Veronika Thost Anni-Yasmin Turhan

Abstract

Rough Description Logics have recently been studied as a means for representing and reasoning with imprecise knowledge. Real-world applications need to exploit reasoning over such knowledge in an efficient way. We describe how the combined approach to query answering can be extended to the rough setting. In particular, we extend both the canonical model and the rewriting procedure such that rough queries over rough $\mathcal{EL}$ ontologies can be answered by considering this information alone.

1 Introduction

One of the main challenges in knowledge representation and reasoning is still to cope with vague and imprecise information in an adequate manner. In the presence of instance data, the reasoning task answering conjunctive queries has become well-investigated over the last years. In this report we investigate answering of conjunctive queries for a variant of the description logic $\mathcal{EL}$ that is capable of expressing imprecise information. Imprecision is found in many knowledge domains, particularly those related to medicine and life sciences. A typical source of imprecision in these domains arises from the level of detail in which the knowledge is described. For example, a disease is usually diagnosed by a series of symptoms that a patient presents, but two individuals, say Ana and Bob, showing the same symptoms might in fact suffer from different maladies. Thus, while these individuals might be *equivalent* from a symptomatic point of view, they might be classified into different illness classes.

One of the many approaches suggested for handling imprecise knowledge is based on *rough* approximations. Unlike fuzzy sets, which allow for arbitrary degrees of membership, rough sets allow for one degree for 'vague' membership, one for definitive membership and one for non-membership. The core idea is to partition the elements in a domain into equivalence classes. This partition is induced by their indiscernibility according to the level of detail currently modeled. An individual belongs to the *upper approximation* of the class C (denoted $\overline{C}$) if it is indiscernible from some element of C. For example, Ana and Bob are in the same *symptomatic equivalence* class. If BOB is diagnosed with, say the Cooties, then ANA potentially has the Cooties, too. In rough terminology, ANA is in the upper approximation of Cooties ($\overline{\mathsf{Cooties}}$). An analogous lower approximation of a class can be defined, too. Intuitively, $\underline{C}$ contains the prototypical elements of the class C: if an element x belongs to $\underline{C}$, then every element indiscernible from x is guaranteed to belong to C.

Rough extensions of Description Logics (DLs) [BCM+07] have been proposed as a formalism capable of expressing and reason over these upper and lower approximations [SKP07]. An example is the rough DL $\mathcal{EL}_\rho$, which extends $\mathcal{EL}$ with two new rough concept constructors: one for the lower and one for the upper approximation. This description logic is investigated in this report. The semantics of this logic is based on interpretations $\mathcal{I}$ that, in addition to the classical interpretation function, define an equivalence relation $\rho^\mathcal{I}$ over the domain elements of $\mathcal{I}$. It has been shown that standard reasoning, such as subsumption or instance checking is decidable in this logic in polynomial time [PZ13]. Intuitively, the idea is to construct a minimal model, called the *canonical model*, that describes all the standard relations between named individuals and concept names in a compact, and easy to read manner. The computation of this kind of model is the core of reasoning algorithms and in particular conjunctive query answering.

Interestingly, there is a very tight connection between canonical models for $\mathcal{EL}$-ontologies, and those for $\mathcal{EL}_\rho$-ontologies. In $\mathcal{EL}$, the canonical interpretation has a domain element x_C for each (sub)concept appearing in the ontology. This element x_C is a representative for the concept C, and every concept containing this element x_C is guaranteed to be a subsumer of C. In the case of $\mathcal{EL}_\rho$, the canonical interpretation $\mathcal{I}_\mathcal{O}$ of an ontology $\mathcal{O}$ can be understood as a more detailed view into the classical canonical model. While each concept C appearing in the ontology still produces a representative x_C, this representative induces a whole equivalence class $[x_C]_\rho$ of ρ, rather than a single domain element. This equivalence class provides information regarding the upper and lower approximations of the concept C. This intuition is depicted in Figure 1(a), where the equivalence classes are depicted as grey boxes. Here, the (partial) interpretation is a model for the GCI $A \sqsubseteq \overline{C}$, since there is an auxiliary element in the class $[x_A]_{\rho^\mathcal{I}}$ that is indistinguishable from x_A, i.e., related to it via ρ, and that belongs to C.

Canonical interpretations are the main means for answering conjunctive queries w.r.t. classical $\mathcal{ELH}_\bot$-ontologies [LTW09]. Essentially, here a canonical interpretation is extended with representatives of all individual names from the ABox as well. The information encoded in this interpretation then suffices to answer the queries w.r.t. this interpretation only. Unfortunately, a naïve application of this idea would provide erroneous answers to some queries; for example, an interpretation like the one in Figure 1(b) could return (x_A, x_C) to the query $\phi(x, y) = \exists z.r(x, z) \wedge r(y, z)$, although this is not true in all models of the ontology. To avoid this problem, one first *rewrites* the query into a first-order query,

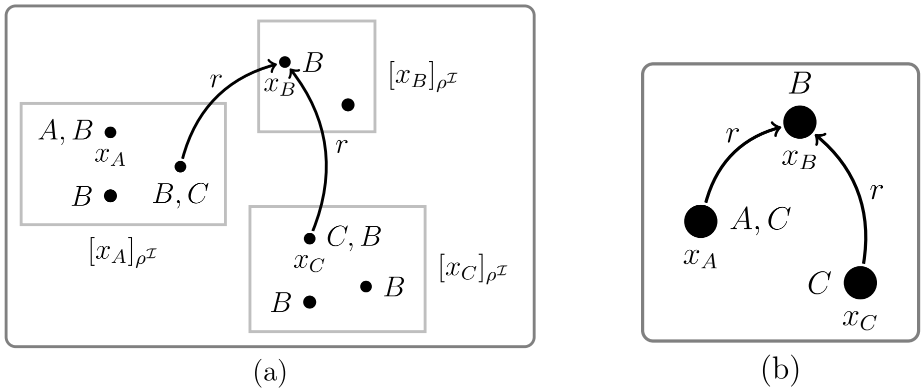


Figure 1: (Partial) canonical interpretations for an $\mathcal{EL}_\rho$- (a) and an $\mathcal{EL}$-ontology (b).

which is then answered over the canonical interpretation. This is known as the *combined approach* [LTW09]. We extend the combined approach for conjunctive query answering in $\mathcal{ELH}_{\bot\rho}$ based on its canonical models.

Since $\mathcal{ELH}_{\bot\rho}$ is an extension of $\mathcal{ELH}_\bot$, all the rewriting rules for query answering in $\mathcal{ELH}_\bot$ apply also in the rough setting. However, the structure of the canonical model of an $\mathcal{ELH}_{\bot\rho}$-ontology is more complex: each symbol gets a representative equivalence class, which is needed to convey the rough approximations of the concepts. Thus, some elements are connected by an equivalence relation, that essentially is a symmetric, transitive and reflexive role ρ. This special kind of role needs to be treated carefully to avoid erroneous answers to a query. Suppose for example, we have an ABox stating that individual a belongs to concept A and that individual c belongs to concept C. We want to answer the query

$$\phi(x_1, x_2) = \exists y_1, y_2.r(x_1, y_1) \wedge r(x_2, y_2) \wedge \rho(y_1, y_2).$$

Here, since ρ is reflexive, the canonical interpretation would, as above, return (x_A, x_C) as an answer. It is thus important to adapt the rewriting technique such that the equivalence relation that the rough constructors yield is handled correctly.

In this report, we describe our extension of the combined approach for computing certain answers to conjunctive queries in the rough DL $\mathcal{ELH}_{\bot\rho}$. As in the case of $\mathcal{ELH}_\bot$-ontologies, the approach consists in computing the canonical interpretation $\mathcal{I}_\mathcal{O}$ that represents all models of the input ontology $\mathcal{O}$, which can be done in polynomial time [PZ13]. This interpretation is used first, as a guide for rewriting a conjunctive query ϕ into a first-order query $\phi^\dagger$, and then as the finite domain over which $\phi^\dagger$ is answered. As a result, we obtain an effective method for answering queries that can allow to model imprecision by rough approximations of a concept—in the ontology as well as in the query.

The report is structured as follows. After defining the syntax and semantics of

$\mathcal{ELH}_{\bot\rho}$ and the reasoning problem studied, query answering, in Section 2, we give the construction of the canonical model in Section 3. The the rewriting is defined in Section 4, and Section 5 finally concludes the report.

2 Preliminaries

In this section, we define the syntax and semantics of $\mathcal{ELH}_{\bot\rho}$, which extends $\mathcal{ELH}$ by the bottom concept $\bot$ and by concept constructors for the *lower approximation* and the *upper approximation*. We then define the problem of answering conjunctive queries in this logic. Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be non-empty, pairwise disjoint sets of *concept names*, *role names*, and *individual names*.

Definition 2.1 ($\mathcal{ELH}_{\bot\rho}$ Syntax)**.** $\mathcal{ELH}_{\bot\rho}$-concepts *are built from concept names $A \in \mathsf{N_C}$ and role names $r \in \mathsf{N_R}$. If C_1 and C_2 are $\mathcal{ELH}_{\bot\rho}$-concepts, then expressions built according to the following syntax rule:*

$$C ::= A \mid \top \mid \bot \mid C_1 \sqcap C_2 \mid \exists r.C_1 \mid \overline{C_1} \mid \underline{C_1}$$

are $\mathcal{ELH}_{\bot\rho}$-concepts as well. Concepts of the form $\overline{C}$ are called upper approximation of C *and concepts of the form $\underline{C}$ are called* lower approximation of C.

The semantics of $\mathcal{ELH}_{\bot\rho}$ is given by interpretations. Here we need to take into account the upper and lower approximation, which is based on the indiscernibility relation ρ. We require that ρ is not an element of the set of role names $\mathsf{N_R}$ and consequently does not appear in $\mathcal{ELH}_{\bot\rho}$-concepts. The main difference between ρ and role names is the fact that ρ is always interpreted as an equivalence relation. Given an interpretation $\mathcal{I}$, $[x]_{\rho^\mathcal{I}}$ denotes the *equivalence class* of an element $x \in \Delta^\mathcal{I}$ w.r.t. the relation $\rho^\mathcal{I}$.

Definition 2.2 (Semantics of $\mathcal{ELH}_{\bot\rho}$-concepts)**.** *A* (rough) interpretation *is a triple $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I}, \rho^\mathcal{I})$, where*

- *the* domain $\Delta^\mathcal{I}$ *is a non-empty set,*
- *$\cdot^\mathcal{I}$ is a function that assigns to every $A \in \mathsf{N_C}$ a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, to every $r \in \mathsf{N_R}$ a binary relation $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, and*
- *the* indiscernibility relation $\rho^\mathcal{I}$ *is an equivalence relation on $\Delta^\mathcal{I}$.*

The function $\cdot^\mathcal{I}$ maps $\top^\mathcal{I} := \Delta^\mathcal{I}$ and $\bot^\mathcal{I} := \emptyset$. It is extended to complex $\mathcal{ELH}_{\bot\rho}$-concepts as follows:

$$\begin{aligned}
(C_1 \sqcap C_2)^\mathcal{I} &:= C_1^\mathcal{I} \cap C_2^\mathcal{I};\\
(\exists r.C)^\mathcal{I} &:= \{x \in \Delta^\mathcal{I} \mid \exists y \in \Delta^\mathcal{I}, (x,y) \in r^\mathcal{I}, y \in C^\mathcal{I}\};\\
\overline{C}^\mathcal{I} &:= \{x \in \Delta^\mathcal{I} \mid [x]^\mathcal{I}_{\rho^\mathcal{I}} \cap C^\mathcal{I} \neq \emptyset\};\\
\underline{C}^\mathcal{I} &:= \{x \in \Delta^\mathcal{I} \mid [x]^\mathcal{I}_{\rho^\mathcal{I}} \subseteq C^\mathcal{I}\}.
\end{aligned}$$

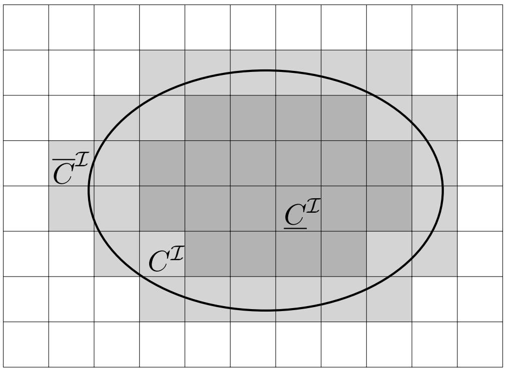


Figure 2: Semantics of a concept, its upper (dark grey) and lower (light grey) approximation.

Intuitively, the indiscernibility relation ρ groups the elements of the domain that cannot be distinguished from each other. The upper approximation $\overline{C}$ of a given concept C describes those elements that cannot be excluded from belonging to C, as they are indistinguishable from some element belonging to this concept. Dually, the individuals $\underline{C}$ are those that are discernible (i.e., can be detached) from each element *not* belonging to C. The extension of a concept in relation to its upper and lower approximation is depicted in Figure 2.

Now, as usual, concepts can be used to build DL ontologies. The terminological component of the ontology is defined as follows.

Definition 2.3 (GCI, RIA, TBox)**.** *Let C and D be $\mathcal{ELH}_{\bot\rho}$-concepts and $r, s \in \mathsf{N_R}$. A* general concept inclusion (GCI) *is an expression of the form $C \sqsubseteq D$, a* role inclusion axiom (RIA) *is an expression of the form $r \sqsubseteq s$. A* TBox *$\mathcal{T}$ is a finite set of GCIs and RIAs.*

$\mathcal{I}$ satisfies *a GCI $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ and a RIA $r \sqsubseteq s$ if $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$. An interpretation that satisfies all GCIs and all RIAs contained in a TBox $\mathcal{T}$ is a* model of the TBox *$\mathcal{T}$.*

Observe, that ρ does neither appear in GCIs nor RIAs. The assertional component of a DL ontology allows to specify facts about objects. Here, in contrast to the TBox, the indiscernibility relation can be used directly.

Definition 2.4 (Assertion, ABox)**.** *Let C be an $\mathcal{ELH}_{\bot\rho}$-concept, $r \in \mathsf{N_R} \cup \{\rho\}$, and $a, b \in \mathsf{N_I}$. A* concept assertion *is an expression of the form $C(a)$ and a* role assertion *is an expression of the form $r(a, b)$. An* ABox *$\mathcal{A}$ is a finite set of assertions. Together, a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$ form an* ontology *$\mathcal{O} = (\mathcal{T}, \mathcal{A})$.*

$\mathcal{I}$ satisfies *a concept assertion $C(a)$ if $a^{\mathcal{I}} \in C^{\mathcal{I}}$, a role assertion $r(a, b), r \in \mathsf{N_R}$, if $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$, and an assertion $\rho(a, b)$ if $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in \rho^{\mathcal{I}}$. An interpretation that satisfies all assertions contained in an ABox $\mathcal{A}$ is a* model of the ABox *$\mathcal{A}$. $\mathcal{I}$ is a* model of an ontology *$\mathcal{O} = (\mathcal{T}, \mathcal{A})$, if it is a model for $\mathcal{T}$ and $\mathcal{A}$.*

We use the standard assumption made for DL systems that all interpretations satisfy the *unique name assumption* (UNA) which means that, for all distinct individual names $a, b \in \mathsf{N_I}$ occurring in α and $\mathcal{A}$, we have $a^\mathcal{I} \neq b^\mathcal{I}$.

Based on the semantics, reasoning services can be defined for ontologies. If it has a model, an ontology is *consistent.* For an axiom, a set of axioms, or an ontology α, we write $\mathcal{I} \models \alpha$, if $\mathcal{I}$ satisfies α. For an ontology $\mathcal{O}$ together with an axiom or a set of axioms α, we further write $\mathcal{O} \models \alpha$, if every model of $\mathcal{O}$ satisfies α.

The reasoning service addressed in this report is answering of conjunctive queries. As customary, we characterize conjunctive queries by means of first order (FO) queries. In this context $\mathsf{N_C}$ and $\mathsf{N_R} \cup \{\rho\}$ are considered as sets of unary and binary FO predicates, respectively. In addition, the indiscernibility relation ρ can be characterized as an equivalence relation.

Definition 2.5 (Syntax of conjunctive queries in $\mathcal{ELH}_{\perp\rho}$)**.** *Let* $\mathsf{N_V}$ *be a set of variables. The elements of* $\mathsf{N_V} \cup \mathsf{N_I}$ *are called* terms. *A* first-order (FO) query *is an FO formula* ϕ *built from terms and the predicates in* $\mathsf{N_C}$ *and* $\mathsf{N_R}$ *.*

We sometimes denote such a query by $\phi(\vec{x})$*, where* $\vec{x} = x_1, \ldots, x_k$ *and* $x_i \in \mathsf{N_V}$ *for* $1 \leq i \leq k$ *are the* free variables *in* ϕ*, which are also called* answer variables *of* $\phi(\vec{x})$*. We call the query* k*-ary, if there are* k *answer variables. The variables occurring in* $\phi(\vec{x})$*, but not in* $\vec{x}$ *are called* quantified variables.

Let C *be an* $\mathcal{ELH}_{\perp\rho}$*-concept,* $r \in \mathsf{N_R} \cup \{\rho\}$ *a role or the indiscernibility relation, and* $t, t' \in \mathsf{N_V} \cup \mathsf{N_I}$*. An* atom *can be a* $\mathcal{ELH}_{\perp\rho}$-concept atom *of the form* $C(t)$ *or a* role atom *of the form* $r(t, t')$*. A* conjunctive query *(CQ) is a FO query of the form* $\phi(\vec{x}) = \exists \vec{y}.\psi(\vec{x}, \vec{y})$*, where* $\vec{y} = y_1, \ldots, y_m \in \mathsf{N_V}$ *and* ψ *is a (possibly empty) finite conjunction of* atoms*. The empty conjunction is denoted by* true*.*

To conveniently access parts of a conjunctive query, we introduce a bit of notation. We denote by

- $\mathsf{Ind}(\phi)$ the set of individuals occurring in a query ϕ,
- $\mathsf{Term}(\phi)$ the set of terms occurring in ϕ,
- $\mathsf{Var}(\phi)$ the set of variables occurring in ϕ,
- $\mathsf{AVar}(\phi)$ the set of answer variables in ϕ, and by
- $\mathsf{QVar}(\phi)$ the set of quantified variables in ϕ.

Note that we sometimes consider a conjunctive query ϕ as the set of atoms occurring in it.

Definition 2.6. *Let* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I}, \rho^\mathcal{I})$ *be an interpretation. A* match *for* $\mathcal{I}$ *and a CQ* ϕ *is a mapping* $\pi \colon \mathsf{Term}(\phi) \to \Delta^\mathcal{I}$ *such that* $\pi(a) = a^\mathcal{I}$ *for all* $a \in \mathsf{Term}(\phi) \cap \mathsf{N_I}$ *and all atoms in* ϕ *are satisfied.*

For a quantifier-free FO query ϕ, the relation $\mathcal{I} \models^\pi \phi$ is defined by induction on the structure of ϕ, as follows:

$$\begin{array}{lll}
\mathcal{I} \models^\pi C(t) & \textit{iff} & \pi(t) \in C^\mathcal{I} \\
\mathcal{I} \models^\pi r(t,t') & \textit{iff} & (\pi(t), \pi(t')) \in r^\mathcal{I} \\
\mathcal{I} \models^\pi \neg\psi & \textit{iff} & \mathcal{I} \not\models^\pi \psi \\
\mathcal{I} \models^\pi \psi_1 \wedge \psi_2 & \textit{iff} & \mathcal{I} \models^\pi \psi_1 \textit{ and } \mathcal{I} \models^\pi \psi_2 \\
\mathcal{I} \models^\pi \psi_1 \vee \psi_2 & \textit{iff} & \mathcal{I} \models^\pi \psi_1 \textit{ or } \mathcal{I} \models^\pi \psi_2
\end{array}$$

In the following, we introduce the central reasoning problem of this paper, namely to compute certain answers to $\mathcal{ELH}_{\bot\rho}$-CQs.

Definition 2.7 (Query Answering)**.** *Let $\phi(\vec{x}) = \exists\vec{y}.\psi(\vec{x},\vec{y})$ be a query with ψ a quantifier-free FO query. If π maps all terms in accordance with $\mathcal{I}$, then a mapping $\pi\colon \mathsf{Term}(\phi) \to \Delta^\mathcal{I}$ is a* match *for ϕ and $\mathcal{I}$ if $\pi(a) = a^\mathcal{I}$ for all $a \in \mathsf{Term}(\phi) \cap \mathsf{N_I}$ and $\mathcal{I} \models^\pi \phi$. Moreover, for $\vec{x} = x_1, \ldots, x_k$ such that $\pi(x_i) = a_i^\mathcal{I}$, $1 \leq i \leq k$, π is called an $(a_1, \ldots, a_k)$-match for $\mathcal{I}$ and ϕ (or* answer *to ϕ w.r.t. $\mathcal{I}$, written $\mathcal{I} \models \phi(a_1, \ldots, a_k)$). Let now ϕ be a k-ary CQ and $\mathcal{O}$ be an ontology. Then, a tuple $(a_1, \ldots, a_k)$, $a_i \in \mathsf{N_I}$ and a_i occurring in $\mathcal{O}$, is a* certain answer *to ϕ w.r.t. $\mathcal{O}$ if $\mathcal{I} \models \phi(a_1, \ldots, a_k)$ holds for every $\mathcal{I}$ with $\mathcal{I} \models \mathcal{O}$.*

The set of all certain answers to ϕ w.r.t. $\mathcal{O}$ is denoted by $\mathsf{Cert}(\phi, \mathcal{O})$.

Since our approach is based on the combined approach by rewriting described in [LTW09], we also use the assumptions made there. So, in the remainder of this report we assume

1. queries to contain only individual names that occur in the ontology they refer to,
2. there are no $r, s \in \mathsf{N_R}$ such that $r \neq s$, $\mathcal{O} \models r \sqsubseteq s$, and $\mathcal{O} \models s \sqsubseteq r$, and
3. $\mathcal{A}$ and ϕ contain only primitive rough concepts.

Note that these assumptions do not represent restrictions since additional individual names can be easily introduced in an ontology by adding tautological assertions to the latter. Moreover, Assumption 2 is satisfied by any ontology if, for example, s is substituted by r in that ontology and the corresponding queries. Assumption 3 is no restriction, since any complex $\mathcal{ELH}_{\bot\rho}$-concept C occurring in $\mathcal{A}$ and ϕ can equivalently be replaced by a fresh concept name A if $A \equiv C$ is added to $\mathcal{T}$.

3 On Canonical Interpretations

The combined approach for answering CQs over an $\mathcal{ELH}_\bot$-ontology $\mathcal{O}$ heavily relies on the so-called *canonical model* $\mathcal{I}_\mathcal{O}$ of $\mathcal{O}$, which represents a materialization of the knowledge encoded in the TBox. In particular, all certain answers to a CQ w.r.t. $\mathcal{O}$ can be retrieved by considering the so-called *unraveling* $\mathcal{U}_\mathcal{O}$ of $\mathcal{I}_\mathcal{O}$. In this section, we show that such interpretations $\mathcal{I}_\mathcal{O}$ (in Section 3.1) and $\mathcal{U}_\mathcal{O}$ (in Section 3.2) can be constructed also in our rough setting. However, due to the presence of the indiscernibility relation and its special semantics, these constructions are more involved. In [PZ13] a completion-based algorithm was given that produces the canonical models for a more expressive DL than $\mathcal{ELH}_{\bot\rho}$. In the following we give a direct definition for anonical models for $\mathcal{ELH}_{\bot\rho}$-ontologies.

3.1 Finite Canonical Interpretations for $\mathcal{ELH}_{\bot\rho}$

This section describes the canonical model $\mathcal{I}_\mathcal{O}$ of the $\mathcal{ELH}_{\bot\rho}$-ontology $\mathcal{O}$ in detail:

1. An introductory example first gives an intuition of our construction.
2. After the formal definition of $\mathcal{I}_\mathcal{O}$, we further adapt it to an interpretation $\mathcal{I}^r_\mathcal{O}$.
3. We show that all the equivalence classes of $\rho^{\mathcal{I}^r_\mathcal{O}}$ are of a special shape.
4. Finally, we show that $\mathcal{I}^r_\mathcal{O}$ can be used to retrieve the certain answers to instance queries, which are a simple form of CQs, and that $\mathcal{I}^r_\mathcal{O}$ is indeed a model of $\mathcal{O}$.

We define a canonical interpretation that describes all the basic relations between symbols in the signature of $\mathcal{O}$ that are entailed by this ontology; the construction is an extension of the canonical models given in [LTW09]. In order to do so, the notion of a subconcept is extended to $\mathcal{ELH}_{\bot\rho}$-concepts in the following way:

$$
\begin{aligned}
\mathsf{Sub}(A) &:= \{A\}, \text{ for } A \in \mathsf{N_C} \cup \{\bot, \top\}\\
\mathsf{Sub}(C \sqcap D) &:= \{C \sqcap D\} \cup \mathsf{Sub}(C) \cup \mathsf{Sub}(D),\\
\mathsf{Sub}(\exists r.C) &:= \{\exists r.C\} \cup \mathsf{Sub}(C),\\
\mathsf{Sub}(\overline{C}) &:= \{\overline{C}\} \cup \mathsf{Sub}(C),\\
\mathsf{Sub}(\underline{C}) &:= \{\underline{C}\} \cup \mathsf{Sub}(C).
\end{aligned}
$$

In what follows, we use $\mathsf{Sub}(\mathcal{T})$ to denote the set of all subconcepts of concepts that occur in GCIs contained in $\mathcal{T}$[1] and $\mathsf{Ind}(\mathcal{A})$ for the set of individual names that occur in $\mathcal{A}$. As in the case of $\mathcal{ELH}_\bot$ canonical models, we use an auxiliary set in which all subconcepts of $\mathcal{T}$ are collected: $\mathsf{N}_\mathsf{I}^{\mathsf{aux}} := \{x_C \mid C \in \mathsf{Sub}(\mathcal{T})\}$.

[1] Observe that $\mathsf{Sub}(\mathcal{T})$ contains all subconcepts in $\mathcal{O}$, since $\mathcal{A}$ only contains concept names.

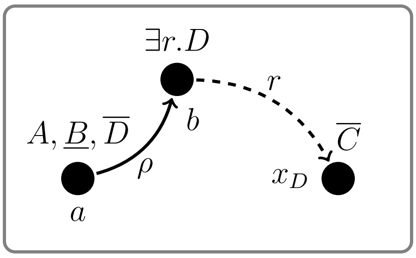


Figure 3: The classical canonical model for the example ontology $\mathcal{O}_{ex}$.

3.1.1 An Example Ontology and its Canonical Model

In this example, we consider $\mathcal{O}_{ex} = (\mathcal{T}_{ex}, \mathcal{A}_{ex})$. Let $a, b \in \mathsf{N_I}, A, B \in \mathsf{N_C}, r \in \mathsf{N_R}$, and

$$\mathcal{T}_{ex} = \{\, C \sqsubseteq A \sqcap \underline{B},\ D \sqsubseteq \overline{C} \,\}$$
$$\mathcal{A}_{ex} = \{\, C(a), \overline{D}(a),\ \exists r.D(b),\ \rho(a,b) \,\}.$$

An illustration of the classical canonical model of $\mathcal{O}_{ex}$ considered as an $\mathcal{ELH}_\bot$-ontology (i.e., without considering the approximations as constructors and considering ρ as an ordinary role, meaning $\underline{B}, \overline{C}, \overline{D} \in \mathsf{N_C}$ and $\rho \in \mathsf{N_R}$) is given in Figure 3. Note that this figure and the following ones show only those elements that are *reachable* from some named individual from $\mathsf{Ind}(\mathcal{A})$.

However, in the rough setting, ρ is an equivalence relation. For our procedure to obtain the canonical model $\mathcal{I}_{\mathcal{O}_{ex}}$, it is hence critical that the equivalence classes of ρ in $\mathcal{I}_{\mathcal{O}_{ex}}$ are defined cautiously. We therefore aim at defining one such class for each named individual and each element in $\mathsf{N_I^{aux}}$ and keep them as separate as long as possible. In particular, the equivalence classes for the $\mathsf{N_I^{aux}}$ elements never merge with other equivalence classes. In contrast to this, ρ-assertions in the ABox, as in $\mathcal{A}_{ex}$, can require the merging of the equivalence classes of named individuals.

To collect all those concepts that are definitely satisfied by all the elements in one equivalence class of $\rho^{\mathcal{I}_\mathcal{O}}$ (created for some element $e \in \mathsf{Ind}(\mathcal{A}) \cup \mathsf{N_I^{aux}}$), we add additional elements of the form ℓ_e. This is depicted in Figure 4, where we still assume $\overline{C}, \overline{D} \in \mathsf{N_C}$, but respect the special semantics of the lower approximation and ρ. This figure also outlines the division of the equivalence classes of $\rho^{\mathcal{I}_\mathcal{O}}$. Note that the borders of the latter are strictly separated by the role edges to elements of $\mathsf{N_I^{aux}}$.

Also note that the figure just depicts the ρ-relations that directly follow from $\mathcal{O}$ meaning without considering the symmetric and transitive closure. Further note that we especially have $\ell_a \in B^{\mathcal{I}_\mathcal{O}}$, because of $a \in \underline{B}^{\mathcal{I}_\mathcal{O}}$. Based on the semantics of ρ, we thus have that all elements in the equivalence class $[a]_{\rho^{\mathcal{I}_\mathcal{O}}} = \{a, \ell_a, b, \ell_b\}$ satisfy B, too.

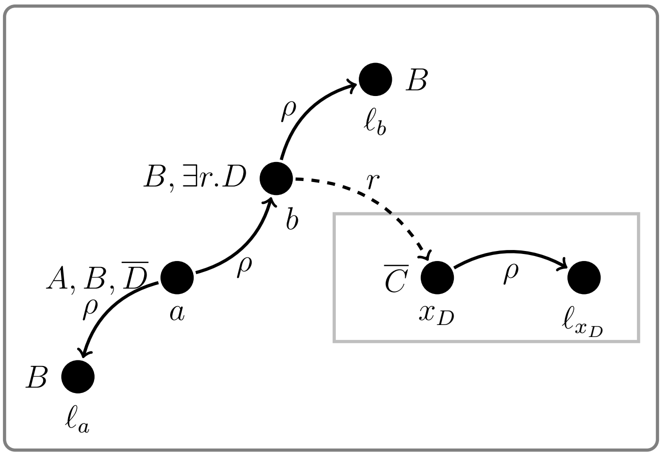


Figure 4: The classical canonical model with the extensions for the lower approximation-constructor (before taking the symmetric and transitive closure to get the full relation $\rho^{\mathcal{I}_\mathcal{O}}$).

To resolve also upper approximation concepts of the form $\overline{C}$, we use additional elements of the form $x_{C,e}$ in the respective equivalence class (i.e., that of e) of $\rho^{\mathcal{I}_\mathcal{O}}$, as it is illustrated in Figure 5. From $x_D \in D^{\mathcal{I}_\mathcal{O}}$ and by $\mathcal{T}_{ex}$, we get $x_D \in \overline{C}^{\mathcal{I}_\mathcal{O}}$, and hence we add $(x_D, x_{C,x_D}) \in \rho^{\mathcal{I}_\mathcal{O}}$. Note that, by resolving the upper approximation, especially for concept $\overline{C}$, we also get that all elements in the concerned equivalence class satisfy B.

The cases exemplified here give an intuition why the construction of the canonial model for $\mathcal{ELH}_{\perp\rho}$ is a little more involved. We now proceed with the formal definition of the canonical interpretation.

3.1.2 The Definition of the Canonical Model $\mathcal{I}_\mathcal{O}$

To ease presentation we assume in the remainder of this section that $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ is an arbitrary, but fixed consistent $\mathcal{ELH}_{\perp\rho}$-ontology with $\mathcal{R}$ the set of RIAs in $\mathcal{T}$, and that ϕ is a CQ which is to be answered w.r.t. $\mathcal{O}$.

To distinguish the different kinds of elements in the domain of $\mathcal{I}_\mathcal{O}$, we use the auxiliary sets $\mathsf{N}_\mathsf{I}^\mathsf{aux}$, $\mathsf{N}_\mathsf{I}^\mathsf{low}$, $\mathsf{N}_\mathsf{I}^\mathsf{up}$, and $\mathsf{N}_\mathsf{I}^\rho$, which are disjoint to N_I:

$$
\begin{aligned}
\mathsf{N}_\mathsf{I}^\mathsf{aux} &:= \{x_C \mid C \in \mathsf{Sub}(\mathcal{T})\},\\
\mathsf{N}_\mathsf{I}^\mathsf{low} &:= \{\ell_e \mid e \in \mathsf{Ind}(\mathcal{A}) \cup \mathsf{N}_\mathsf{I}^\mathsf{aux}\},\\
\mathsf{N}_\mathsf{I}^\mathsf{up} &:= \{x_{C,e} \mid C \in \mathsf{Sub}(\mathcal{T}), e \in \mathsf{Ind}(\mathcal{A}) \cup \mathsf{N}_\mathsf{I}^\mathsf{aux}\}, \text{ and}\\
\mathsf{N}_\mathsf{I}^\rho &:= \mathsf{N}_\mathsf{I}^\mathsf{low} \cup \mathsf{N}_\mathsf{I}^\mathsf{up}.
\end{aligned}
$$

Intuitively, the elements from these sets represent all the different sets of concepts that need to be distinguished by $\mathcal{I}_\mathcal{O}$ in order to satisfy $\mathcal{O}$, as it was already outlined

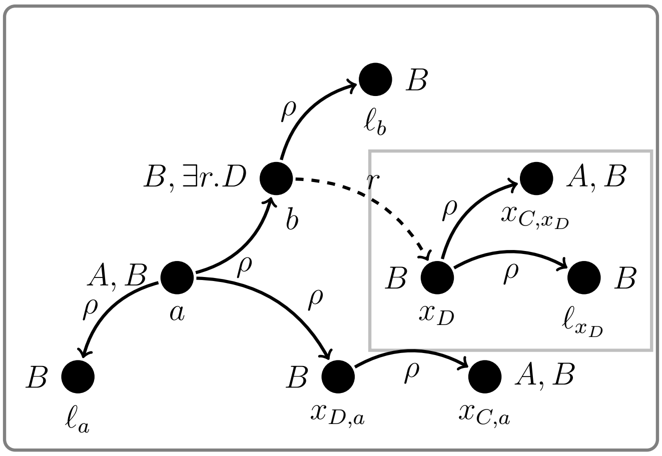


Figure 5: The classical canonical model with the rough extensions.

in the above example.

- $x_C \in \mathsf{N}_\mathsf{I}^{\mathsf{aux}}$ is a canonical role-successor for the TBox (sub-)concept C and from N_R.
- $\ell_e \in \mathsf{N}_\mathsf{I}^{\mathsf{low}}$ is in the same equivalence class as e (i.e., $\rho(\ell_e, e)$ holds)). The element ℓ_e represents the set of those concepts of which all the elements in this equivalence class $[e]$ are in the lower approximations.
- $x_{C,e} \in \mathsf{N}_\mathsf{I}^{\mathsf{up}}$ is a representative for those elements that are indiscernible from e and satisfy the TBox (sub-)concept C, i.e. $\rho(x_{C,e}, e)$ holds and $x_{C,e}$ is an instance of C. Thus $x_{C,e}$ is in the upper approximation of C; and
- $\mathsf{N}_\mathsf{I}^{\rho}$ collects the auxiliary elements that are representatives for upper or lower approximations.

Using these auxiliary sets, we define the canonical interpretation as follows.

Definition 3.1 (Canonical Interpretation)**.** *The* canonical interpretation *of an ontology $\mathcal{O}$ (with the indiscernibility relation ρ) is defined as $\mathcal{I}_\mathcal{O} = (\Delta^{\mathcal{I}_\mathcal{O}}, \cdot^{\mathcal{I}_\mathcal{O}}, \rho^{\mathcal{I}_\mathcal{O}})$, where*

- $\Delta^{\mathcal{I}_\mathcal{O}} := \mathsf{Ind}(\mathcal{A}) \cup \mathsf{N}_\mathsf{I}^{\mathsf{aux}} \cup \mathsf{N}_\mathsf{I}^{\rho}$*;*
- *for all $a \in \mathsf{Ind}(\mathcal{A})$, $a^{\mathcal{I}_\mathcal{O}} := a$;*

- *for all* $A \in \mathsf{N_C}$

$$
\begin{aligned}
A^{\mathcal{I}_\mathcal{O}} := \ & \{a \in \mathsf{Ind}(\mathcal{A}) \mid \mathcal{O} \models A(a)\} \cup \\
& \{x_C \in \mathsf{N_I^{aux}} \mid \mathcal{O} \models C \sqsubseteq A\} \cup \\
& \{x_{C,e} \in \mathsf{N_I^{up}} \mid \mathcal{O} \models C \sqsubseteq A\} \cup \\
& \{x_{C,b}, \ell_b \in \mathsf{N_I^{\rho}} \mid \mathcal{O} \models \underline{A}(b)\} \cup \\
& \{x_{C,x_D}, \ell_{x_D} \in \mathsf{N_I^{\rho}} \mid \mathcal{O} \models D \sqsubseteq \underline{A}\};
\end{aligned}
$$

- *for all* $r \in \mathsf{N_R}$

$$
\begin{aligned}
r^{\mathcal{I}_\mathcal{O}} := \ & \{(a,b) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mid s(a,b) \in \mathcal{A}, \mathcal{O} \models s \sqsubseteq r\} \cup \\
& \{(a, x_C) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{N_I^{aux}} \mid \mathcal{O} \models \exists r.C(a)\} \cup \\
& \{(x_C, x_D) \in \mathsf{N_I^{aux}} \times \mathsf{N_I^{aux}} \mid \mathcal{O} \models C \sqsubseteq \exists r.D\} \cup \\
& \{(x_{C,e}, x_D) \in \mathsf{N_I^{up}} \times \mathsf{N_I^{aux}} \mid \mathcal{O} \models C \sqsubseteq \exists r.D\} \cup \\
& \{(x_{C,b}, x_D), (\ell_b, x_D) \in \mathsf{N_I^{\rho}} \times \mathsf{N_I^{aux}} \mid \mathcal{O} \models \underline{\exists r.D}(b)\} \cup \\
& \{(x_{C,x_E}, x_D), (\ell_{x_E}, x_D) \in \mathsf{N_I^{\rho}} \times \mathsf{N_I^{aux}} \mid \mathcal{O} \models E \sqsubseteq \underline{\exists r.D}\};
\end{aligned}
$$

- $\rho^{\mathcal{I}_\mathcal{O}}$ *is based on the relation:*

$$
\begin{aligned}
\rho_\mathcal{O} := \ & \{(a,b) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mid \rho(a,b) \in \mathcal{A}\} \cup \\
& \{(a, x_{C,a}) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{N_I^{up}} \mid \mathcal{O} \models \overline{C}(a)\} \cup \\
& \{(x_C, x_{D,x_C}) \in \mathsf{N_I^{aux}} \times \mathsf{N_I^{up}} \mid \mathcal{O} \models C \sqsubseteq \overline{D}\} \cup \\
& \{(x_{C,e}, x_{D,e}) \in \mathsf{N_I^{\rho}} \times \mathsf{N_I^{\rho}} \mid \mathcal{O} \models C \sqsubseteq \overline{D}\} \cup \bigcup_{\ell_e \in \mathsf{N_I^{low}}} \{(e, \ell_e)\}.
\end{aligned}
$$

We define $\rho^{\mathcal{I}_\mathcal{O}}$ *to be the reflexive, symmetric, and transitive closure of* $\rho_\mathcal{O}$*:* $\rho^{\mathcal{I}_\mathcal{O}} := \big(\rho_\mathcal{O} \ \cup \ \{(e', e) \mid (e, e') \in \rho_\mathcal{O}\}\big)^*$.

Note that this definition of $\mathcal{I}_\mathcal{O}$ extends the standard notion of a canonical model in $\mathcal{EL}$ as proposed in the literature (e.g., in [LTW09]). The extension is required to handle the upper and lower approximations introduced by the rough constructors, and it is realized by the new elements in $\mathsf{N_I^{\rho}}$ added to the domain. Moreover, the semantics of $\mathcal{ELH}_{\bot\rho}$ requires ρ to be extended to an equivalence relation $\rho^{\mathcal{I}_\mathcal{O}}$ over the elements of $\Delta^{\mathcal{I}_\mathcal{O}}$. Nevertheless, the cardinality of $\Delta^{\mathcal{I}_\mathcal{O}}$ is polynomial in the size of $\mathcal{O}$. In addition, $\mathcal{I}_\mathcal{O}$ can be computed in polynomial time [PZ13] and also consistency of $\mathcal{O}$ can be checked in polynomial time [PZ13].

3.1.3 About $\rho^{\mathcal{I}_\mathcal{O}^r}$

As described before, the scope of the elements in $\mathsf{N_I^{\rho}}$ is to describe all possible kinds of elements that are indiscernible from those in $\mathsf{N_I^{aux}}$ and $\mathsf{Ind}(\mathcal{A})$. In particular, and as it is stated by the following proposition, different elements in $\mathsf{N}_\mathsf{I}^{\mathsf{aux}}$ are never related via $\rho^{\mathcal{I}_\mathcal{O}}$; moreover, elements from $\mathsf{Ind}(\mathcal{A})$ can only be related in very specific cases. The following proposition follows directly from the definition of $\rho^{\mathcal{I}_\mathcal{O}}$.

Proposition 3.2. *Let $a \in \mathsf{Ind}(\mathcal{A})$ and $x_C \in \mathsf{N}_\mathsf{I}^{\mathsf{aux}}$. Then, for every element $e \in \Delta^{\mathcal{I}_\mathcal{O}}$, the following holds:*

- *if $e \in [a]_{\rho^{\mathcal{I}_\mathcal{O}}}$, then either $e \in \mathsf{Ind}(\mathcal{A})$ or e is of the form $x_{C,b}$ or ℓ_b for some $b \in \mathsf{Ind}(\mathcal{A})$; and*
- *if $e \in [x_C]_{\rho^{\mathcal{I}_\mathcal{O}}}$, then either $e = x_C$ or e is of the form x_{D,x_C} or ℓ_{x_C}.*

To be able to use $\mathcal{I}_\mathcal{O}$ for answering (even instance) queries, we have to make sure that we do not have unnecessary elements in our interpretation. Otherwise, for example, a query $\phi = \exists y.D(y)$ w.r.t. an ontology $\mathcal{O} = (\{C \sqsubseteq D\}, \emptyset)$ would yield true as answer in $\mathcal{I}_\mathcal{O}$, which clearly is no certain answer to the query.

We therefore restrict $\mathcal{I}_\mathcal{O}$ to the elements that are reachable from the individuals $a \in \mathsf{Ind}(\mathcal{A})$. A *path* in $\mathcal{I}_\mathcal{O}$ is a finite sequence $d_0 r_1 d_1 \cdots r_n d_n$, $n \geq 0$, such that $d_0 \in \mathsf{Ind}(\mathcal{A})$, $d_j \in \Delta^{\mathcal{I}_\mathcal{O}} \setminus \mathsf{Ind}(\mathcal{A})$ for all $j > 0$, $r_i \in \mathsf{N}_\mathsf{R} \cup \{\rho_\mathcal{O}\}$ and $(d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}_\mathcal{O}}$ for all $i < n$. We denote the set of all paths in $\mathcal{I}_\mathcal{O}$ as $\mathsf{Paths}(\mathcal{I}_\mathcal{O})$ and the last element d_n in a path $p = d_0 r_1 d_1 \cdots r_n d_n$ as $\mathsf{Tail}(p)$. The interpretation $\mathcal{I}_\mathcal{O}^r$ is obtained by restricting the domain of $\mathcal{I}_\mathcal{O}$ to the set $\{\mathsf{Tail}(p) \mid p \in \mathsf{Paths}(\mathcal{I}_\mathcal{O})\}$.

Notice that in the definition of the paths, we consider only elements that are reachable through the relation $\rho_\mathcal{O}$, and not through its closure $\rho^{\mathcal{I}_\mathcal{O}}$. It can be easily seen that every element that is reachable from an individual name through roles and the relation $\rho^{\mathcal{I}_\mathcal{O}}$ in $\mathcal{I}_\mathcal{O}$ is also reachable through a path in $\mathsf{Paths}(\mathcal{I}_\mathcal{O})$. Thus, $\mathcal{I}_\mathcal{O}^r$ contains all the reachable elements.

Lemma 3.3. *For all $e \in \Delta^{\mathcal{I}_\mathcal{O}^r}$ there is a sequence $d_0, \ldots, d_n \in \Delta^{\mathcal{I}_\mathcal{O}^r}$ and a sequence $r_0, \ldots, r_{n-1} \in \mathsf{N}_\mathsf{R} \cup \{\rho\}$ such that $d_0 \in \mathsf{Ind}(\mathcal{A})^{\mathcal{I}_\mathcal{O}^r}$, $d_n = e$, $(d_i, d_{i+1}) \in r^{\mathcal{I}_\mathcal{O}^r}$ if $r \in \mathsf{N}_\mathsf{R}$ and $(d_i, d_{i+1}) \in \rho_\mathcal{O}$ if $r = \rho$ for all $0 \leq i < n$.*

In the following lemma, we describe some additional properties of the equivalence classes defined by this restricted interpretation.

Lemma 3.4. *Let C, D, and E be arbitrary concepts, and $a, b \in \mathsf{Ind}(\mathcal{A})$.*

(1) if $x_{C,b} \in \Delta^{\mathcal{I}_\mathcal{O}^r}$, then $\{x_{C,b}, \ell_b, b\} \subseteq [b]_{\rho^{\mathcal{I}_\mathcal{O}^r}}$.

(2) $\{x_{C,x_D}, \ell_{x_D}, x_D\} \subseteq [x_E]_{\rho^{\mathcal{I}_\mathcal{O}^r}}$ iff $D = E$ and $x_{C,x_D} \in \Delta^{\mathcal{I}_\mathcal{O}^r}$.

Proof. [(1)] Since $\rho^{\mathcal{I}_\mathcal{O}^r}$ is reflexive, $b \in \mathsf{Ind}(\mathcal{A})$, and $(b, l_b) \in \rho_\mathcal{O}$, we immediately have that $\{b, \ell_b\} \subseteq [b]_{\rho^{\mathcal{I}_\mathcal{O}^r}}$. If $x_{C,b} \in \Delta^{\rho^{\mathcal{I}_\mathcal{O}^r}}$, then there is a $p \in \mathsf{Paths}(\mathcal{I}_\mathcal{O})$ with

$\mathsf{Tail}(p) = x_{C,b}$. Suppose that there is a sequence $d_i r_{i+1} d_{i+1}$ in this path such that $r_{i+1} \in \mathsf{N_R}$. By the latter and the definition of $\mathcal{I}^r_{\mathcal{O}}$, d_{i+1} must be of the form x_D for some concept D, and $\mathsf{Tail}(p)$ cannot be of the form $x_{C,b}$. Thus, p must be of the form $d_0 \rho_{\mathcal{O}} d_1 \rho_{\mathcal{O}} \cdots \rho_{\mathcal{O}} d_n$. Moreover, $d \rho_{\mathcal{O}} x_{D,b}$ can hold only if d is the individual name b, or of the form $x_{E,b} \notin \mathsf{Ind}(\mathcal{A})$. This implies that, for the first element of this path, we have $d_0 = b \in \mathsf{Ind}(\mathcal{A})$, and hence $x_{C,b} \in [b]_{\rho^{\mathcal{I}^r_{\mathcal{O}}}}$.

[(2)] By Proposition 3.2, $[x_E]_{\rho^{\mathcal{I}^r_{\mathcal{O}}}}$ can only contain elements of the form x_E, x_{D,x_E}, or ℓ_{x_E}. Thus, if $\{x_{C,x_D}, \ell_{x_D}, x_D\} \subseteq [x_E]_{\rho^{\mathcal{I}^r_{\mathcal{O}}}}$, for some concept D, then D must be the concept E. For the converse, we can prove analogously to (1) that $\{x_D, \ell_{x_D}\} \subseteq [x_D]_{\rho^{\mathcal{I}^r_{\mathcal{O}}}}$. By the definition of $\mathcal{I}^r_{\mathcal{O}}$, it follows that any path p with $\mathsf{Tail}(p) = x_{C,x_D}$ must contain x_D, and use only the relation $\rho_{\mathcal{O}}$ between x_D and the tail. This then implies that $x_{C,x_D} \in [x_D]_{\rho^{\mathcal{I}^r_{\mathcal{O}}}}$. □

3.1.4 $\mathcal{I}_{\mathcal{O}}$ is a model of $\mathcal{O}$

Having established important properties about $\rho^{\mathcal{I}^r_{\mathcal{O}}}$ We now can show that $\mathcal{I}^r_{\mathcal{O}}$ is a model of $\mathcal{O}$ whenever this ontology is consistent. Moreover, this model provides relevant information about the properties of all models of $\mathcal{O}$, which, among other reasoning tasks, can be used to answer instance queries. We start by showing that several entailments can be obtained from $\mathcal{I}^r_{\mathcal{O}}$, which makes it easy to show that $\mathcal{I}^r_{\mathcal{O}}$ is a model of $\mathcal{O}$, afterwards.

Lemma 3.5. *Let C, D, E be $\mathcal{ELH}_{\bot\rho}$, and $a, b \in \mathsf{Ind}(\mathcal{A})$.*

(1) $a \in C^{\mathcal{I}^r_{\mathcal{O}}}$ iff $\mathcal{O} \models C(a)$

(2) $x_D \in C^{\mathcal{I}^r_{\mathcal{O}}}$ iff $\mathcal{O} \models D \sqsubseteq C$

(3) $x_{D,a} \in C^{\mathcal{I}^r_{\mathcal{O}}}$ iff $\mathcal{O} \models D \sqsubseteq C$ or $\mathcal{O} \models \underline{C}(a)$

(4) $x_{D,x_E} \in C^{\mathcal{I}^r_{\mathcal{O}}}$ iff $\mathcal{O} \models D \sqsubseteq C$ or $\mathcal{O} \models E \sqsubseteq \underline{C}$

(5) $\ell_b \in C^{\mathcal{I}^r_{\mathcal{O}}}$ iff $\mathcal{O} \models \underline{C}(b)$

(6) $\ell_{x_D} \in C^{\mathcal{I}^r_{\mathcal{O}}}$ iff $\mathcal{O} \models D \sqsubseteq \underline{C}$

Proof. We prove the items simultaneously by induction on the structure of C. The base case where $C \in \mathsf{N_C}$ is a direct consequence of the definition of $\mathcal{I}^r_{\mathcal{O}}$. If $C = C_1 \sqcap C_2$, the result follows trivially from the semantics and the induction hypothesis. We now consider the remaining cases in detail.

($C = \exists r.C_1$) ($\Rightarrow$) (1) If $a \in (\exists r.C_1)^{\mathcal{I}^r_\mathcal{O}}$, then there is an $e \in \Delta^{\mathcal{I}^r_\mathcal{O}}$ such that $(a, e) \in r^{\mathcal{I}^r_\mathcal{O}}$ and $e \in C_1^{\mathcal{I}^r_\mathcal{O}}$. By the definition of $\mathcal{I}^r_\mathcal{O}$, $e \notin \mathsf{N}_\mathsf{I}^\rho$. If $e \in \mathsf{N}_\mathsf{I}$, then $s(a, e) \in \mathcal{A}$ for some role s with $\mathcal{O} \models s \sqsubseteq r$. By induction hypothesis, we have that $\mathcal{O} \models C_1(e)$; hence $\mathcal{O} \models \exists r.C_1(a)$. Otherwise, if e is of the form $e = x_D \in \mathsf{N}_\mathsf{I}^{\mathsf{aux}}$, then $\mathcal{O} \models \exists r.D(a)$. Since $\mathcal{O} \models D \sqsubseteq D$, the induction hypothesis yields $x_D \in D^{\mathcal{I}^r_\mathcal{O}}$, and hence $\mathcal{O} \models D \sqsubseteq C_1$. This implies that $\mathcal{O} \models \exists r.C_1(a)$. The remaining items can be treated analogously.

($\Leftarrow$) (1) If $\mathcal{O} \models \exists r.C_1(a)$, then $(a, x_{C_1}) \in r^{\mathcal{I}^r_\mathcal{O}}$, by definition; the induction hypothesis also yields $x_{C_1} \in C_1^{\mathcal{I}^r_\mathcal{O}}$. Hence, $a \in (\exists r.C_1)^{\mathcal{I}^r_\mathcal{O}}$ follows. The proof for the other items is analogous.

($C = \overline{C_1}$) ($\Rightarrow$) (1) If $a \in \overline{C_1}^{\mathcal{I}^r_\mathcal{O}}$, then there is an $e \in \Delta^{\mathcal{I}^r_\mathcal{O}}$ with $(a, e) \in \rho^{\mathcal{I}^r_\mathcal{O}}$ and $e \in C_1^{\mathcal{I}^r_\mathcal{O}}$. By Proposition 3.2, either $e \in \mathsf{N}_\mathsf{I}$, or e is of the form $x_{D,b}$ or ℓ_b for some $b \in \mathsf{Ind}(\mathcal{A})$ and concept D. If $e \in \mathsf{N}_\mathsf{I}$, then by the induction hypothesis, $\mathcal{O} \models C_1(e)$, and hence $\mathcal{O} \models \overline{C_1}(a)$. If e is of the form $x_{D,b}$, we either get $\mathcal{O} \models D \sqsubseteq C_1$ or $\mathcal{O} \models \underline{C_1}(b)$ by IH (3). In the latter case, the semantics directly yields $\mathcal{O} \models \overline{C_1}(a)$ since $(a, e) \in \rho^{\mathcal{I}^r_\mathcal{O}}$. For the former case, $x_{D,b} \in \Delta^{\mathcal{I}^r_\mathcal{O}}$ together with the definition of $\mathcal{I}^r_\mathcal{O}$ implies $\mathcal{O} \models \overline{D}(b)$. Thus, $\mathcal{O} \models \overline{C_1}(b)$. Since $(a, b) \in \rho^{\mathcal{I}^r_\mathcal{O}}$, the semantics yields $\mathcal{O} \models \overline{C_1}(a)$. If e is of the form ℓ_b, Lemma 3.4 (1) yields $(a, b) \in \rho^{\mathcal{I}^r_\mathcal{O}}$. By IH (5), we additionally have $\mathcal{O} \models \underline{C_1}(b)$ and thus $\mathcal{O} \models \overline{C_1}(a)$. The proof for item (2) is very similar. For (3), we can restrict ourselves to the same kinds of elements e as in the proof of item (1), by Proposition 3.2. Then, $x_{D,a} \in \overline{C_1}^{\mathcal{I}^r_\mathcal{O}}$ implies $a \in \overline{C_1}^{\mathcal{I}^r_\mathcal{O}}$. By IH (1), we thus get $\mathcal{O} \models \overline{C_1}(a)$, which corresponds to $\mathcal{O} \models \underline{(\overline{C_1})}(a)$. The proof of (5) is analogous to the one of (3), and the proofs of items (4) and (6) correspond the one of (2) in the same way.

($\Leftarrow$) (1) If $\mathcal{O} \models \overline{C_1}(a)$, then $(a, x_{C_1,a}) \in \rho^{\mathcal{I}^r_\mathcal{O}}$. We then can apply IH (3) to obtain $x_{C_1,a} \in C_1^{\mathcal{I}^r_\mathcal{O}}$. But then, the semantics directly yields $a \in \overline{C_1}^{\mathcal{I}^r_\mathcal{O}}$. The proof for item (2) is analogous. For (3), if $\mathcal{O} \models D \sqsubseteq \overline{C_1}$ holds, the proof is analogous to the one of (1) and (2). Assume $\mathcal{O} \models \underline{(\overline{C_1})}(a)$. We then have $a \in \overline{C_1}^{\mathcal{I}^r_\mathcal{O}}$ by IH (1); the semantics then yields $x_{D,a} \in \overline{C_1}^{\mathcal{I}^r_\mathcal{O}}$. The proof of item (4) is analogous, and the proofs of (5) and (6) are analogous to the second case in the proof of (3) and (4), respectively.

($C = \underline{C_1}$) ($\Rightarrow$) (1) If $a \in \underline{C_1}^{\mathcal{I}^r_\mathcal{O}}$, then all elements that are $\rho^{\mathcal{I}^r_\mathcal{O}}$-related to a satisfy C_1, too. By Lemma 3.4, $(a, \ell_a) \in \rho^{\mathcal{I}^r_\mathcal{O}}$ and hence $\ell_a \in C_1^{\mathcal{I}^r_\mathcal{O}}$. IH (5) directly leads to $\mathcal{O} \models \underline{C_1}(a)$. The proof for the other items is analogous.

($\Leftarrow$) (1) Suppose that $\mathcal{O} \models \underline{C_1}(a)$ and that there is an element $e \in \Delta^{\mathcal{I}^r_\mathcal{O}}$ such that $a\rho^{\mathcal{I}^r_\mathcal{O}}e$ and $e \notin C_1^{\mathcal{I}^r_\mathcal{O}}$. By Proposition 3.2, e is either an individual name or of the form $x_{D,b}$ or ℓ_b for some concept D and $b \in \mathsf{N}_\mathsf{I}$. If $e \in \mathsf{N}_\mathsf{I}$, we have

$\rho(a,b) \in \mathcal{A}$, by Lemma 3.4, and hence get $\mathcal{O} \models C_1(b)$, by the semantics. But then, the application of (IH 1) yields $b \in C_1^{\mathcal{I}^r_\mathcal{O}}$, which is a contradiction. If e is of the form $e = x_{D,b}/a_b$, we have $x_{D,b}/a_b \in [b]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$, by Lemma 3.4(1), and thus get $b \in [a]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$, by the semantics. Lemma 3.4(2) then yields $\rho(a,b) \in \mathcal{A}$. Given $\mathcal{O} \models \underline{C_1}(a)$, the semantics leads to $\mathcal{O} \models \underline{C_1}(b)$. Then, the application of (IH 3/5) yields $\mathcal{O} \not\models \underline{C_1}(b)$, which is a contradiction.

For (3), there are two possible cases to be considered. However, given $x_{D,a} \in \Delta^{\mathcal{I}^r_\mathcal{O}}$ (i.e., it is reachable in $\mathcal{I}^r_\mathcal{O}$), the definition of $\mathcal{I}^r_\mathcal{O}$ yields that $\mathcal{O} \models \overline{D}(a)$. But then, the first case, $\mathcal{O} \models D \subseteq \underline{C_1}$, by the semantics, implies the second case $\mathcal{O} \models \underline{C_1}(a)$. Given $\mathcal{O} \models \underline{C_1}(a)$, the proof basically follows the one of (1) and only differs from the latter in that Lemma 3.4 has to be applied for the case $e = b \in \mathsf{Ind}\mathcal{A}$ to obtain $x_{D,a} \in [a]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$ and get $a\rho^{\mathcal{I}^r_\mathcal{O}}b$, by the transitivity of $\rho^{\mathcal{I}^r_\mathcal{O}}$. Having also the assumption that $\mathcal{O} \models \underline{C_1}(a)$, the proof of (5) corresponds to the one of (3). The proofs of (2), (4), and (6) are similar, but less involved, because the contradicting assumption, by applying the induction hypothesis, always directly yields a contradiction. For example, in the proof of (2), the assumption is $\mathcal{O} \models D \sqsubseteq \underline{C_1}$, and in the case $e = x_{E,x_D} \notin C_1$, the application of (IH 4) yields $\mathcal{O} \not\models D \sqsubseteq C_1$. $\square$

Lemma 3.6. *If $\mathcal{O}$ is consistent, then $\mathcal{I}^r_\mathcal{O}$ is a model of $\mathcal{O}$.*

Proof. By definition, $\mathcal{I}^r_\mathcal{O}$ is a model of $\mathcal{A}$ and all role inclusions in $\mathcal{T}$. Let now $C_1 \sqsubseteq C_2 \in \mathcal{T}$ and $x \in C_1^{\mathcal{I}^r_\mathcal{O}}$. If $x \in \mathsf{N_I}$, Lemma 3.5 (1) yields $\mathcal{O} \models C_1(x)$ and $\mathcal{O} \models C_2(x)$ since $\mathcal{O} \models C_1 \sqsubseteq C_2$ holds. Applying Lemma 3.5 (1) leads to $x \in C_2^{\mathcal{I}^r_\mathcal{O}}$. The cases with $x \in \mathsf{N_I^{aux}}$ and $x \in \mathsf{N_I^{\rho}}$ can be treated analogously. $\square$

When CQ-answering is considered, another problem, which has already been outlined in Section 1, is the reuse of elements of $\mathsf{N_I^{aux}}$ representing the role-successors and consequently also that of the elements from $\mathsf{N_I^{\rho}}$ in $\mathcal{I}^r_\mathcal{O}$ (since they are connected to the $\mathsf{N_I^{aux}}$-elements). To cope with that, we define another interpretation based on $\mathcal{I}^r_\mathcal{O}$, next.

3.2 An Interpretation for Query Answering

Based on the deficiencies of $\mathcal{I}^r_\mathcal{O}$, we now construct the unraveling $\mathcal{U}_\mathcal{O}$ of $\mathcal{I}^r_\mathcal{O}$. In a nutshell, this interpretation is obtained by considering the paths in $\mathcal{I}^r_\mathcal{O}$ as domain elements of $\mathcal{U}_\mathcal{O}$. For later proofs, it is additionally important that we establish a certain correspondence (i.e., a surjective mapping) between the domain elements of $\mathcal{U}_\mathcal{O}$ and those of $\mathcal{I}^r_\mathcal{O}$.

Hence, this section covers the following:

1. We formally define $\mathcal{U}_\mathcal{O}$ and show that its domain elements (i.e., the paths) are not arbitrary, but of some specific structure.

2. We show that there is indeed a mapping as mentioned above, which, in particular, is surjective w.r.t. the relation ρ.

3. Finally, we show that $\mathcal{U}_\mathcal{O}$ is a model of $\mathcal{O}$ and can be used to retrieve the certain answers to CQs w.r.t. $\mathcal{O}$.

3.2.1 The definition of $\mathcal{U}_\mathcal{O}$

We define the interpretation $\mathcal{U}_\mathcal{O} = (\Delta^{\mathcal{U}_\mathcal{O}}, \cdot^{\mathcal{U}_\mathcal{O}}, \rho^{\mathcal{U}_\mathcal{O}})$, called the *unraveling* of $\mathcal{I}^r_\mathcal{O}$, where $\Delta^{\mathcal{U}_\mathcal{O}} := \mathsf{Paths}_\mathcal{A}(\mathcal{I}^r_\mathcal{O})$, and

$$\begin{aligned}
a^{\mathcal{U}_\mathcal{O}} &:= a, \text{ for all } a \in \mathsf{Ind}(\mathcal{A}),\\
A^{\mathcal{U}_\mathcal{O}} &:= \{p \mid \mathsf{Tail}(p) \in A^{\mathcal{I}^r_\mathcal{O}}\}, \text{ for all } A \in \mathsf{N_C},\\
r^{\mathcal{U}_\mathcal{O}} &:= \{(a,b) \mid a,b \in \mathsf{Ind}(\mathcal{A}), (a,b) \in r^{\mathcal{I}^r_\mathcal{O}}\} \cup\\
&\qquad \{(p, p\cdot se) \mid p, p\cdot se \in \Delta^{\mathcal{U}_\mathcal{O}}, \mathcal{R} \models s \sqsubseteq r\}, \text{ for all } r \in \mathsf{N_R}, \text{ and}\\
\rho^{\mathcal{U}_\mathcal{O}} &:= \rho^*_{\mathcal{O}'}, \text{ with}\\
\rho_{\mathcal{O}'} &:= \{(a,b) \mid a,b \in \mathsf{Ind}(\mathcal{A}), (a,b) \in \rho^{\mathcal{I}^r_\mathcal{O}}\} \cup \{(p, p\cdot \rho e) \mid p \cdot \rho e \in \Delta^{\mathcal{U}_\mathcal{O}}\}.
\end{aligned}$$

In this definition $u \cdot v$ denotes the concatenation of u and v. Note that the construction of $\mathcal{U}_\mathcal{O}$ does not depend on the GCIs in $\mathcal{T}$, but only on $\mathcal{R}$.

We now start proving some relevant properties of $\mathcal{U}_\mathcal{O}$. Proposition 3.7 concretizes the kinds of paths that can occur as elements of $\Delta^{\mathcal{U}_\mathcal{O}}$.

Proposition 3.7. *For every $p = d_0 r_1 d_1 \cdots r_n d_n \in \Delta^{\mathcal{U}_\mathcal{O}}$ one of the following conditions hold:*

(i) $p = d_0 \in \mathsf{Ind}(\mathcal{A})$;

(ii) $d_n = x \in \mathsf{N_I^{aux}}$, *and p is of the form $p'rx$ for some $r \in \mathsf{N_R}$ and $p' \in \Delta^{\mathcal{U}_\mathcal{O}}$;*

(iii) $d_n = x_{C,a} \in \mathsf{N_I^{up}}$, $a \in \mathsf{Ind}(\mathcal{A})$, *and p is of the form $a\rho x_{C_1,a} \cdots \rho x_{C_{n-1},a} \rho x_{C,a}$;*

(iv) $d_n = x_{D,x_C} \in \mathsf{N_I^{up}}$, $x_C \in \mathsf{N_I^{aux}}$, *and there is a path p' such that p is of the form $p' r x_C \rho x_{D_1,x_C} \cdots \rho x_{D_i,x_C} \rho x_{D,x_C}$, $i \geq 0$; or*

(v) $d_n = \ell_e \in \mathsf{N_I^{low}}$ *and p is of the form $p' e \rho \ell_e$.*

Proof. Notice that the five conditions consider all possible cases for the last element d_n. Hence, it suffices to show that the type of element used enforces the corresponding shape of the path p. We first consider (i). Since the definition of path states that individuals can only appear in the first position of a path, we must have $\mathsf{Tail}(d) = d_n = d_0$ if $d_n \in \mathsf{Ind}(\mathcal{A})$.

Consider now (ii). If $d_n \in \mathsf{N}_\mathsf{I}^{\mathsf{aux}}$, then by the definition of $\Delta^{\mathcal{U}_\mathcal{O}}$, p must be of the form $p'rx$ for some $p' \in \Delta^{\mathcal{U}_\mathcal{O}}$ and $r \in \mathsf{N}_\mathsf{R} \cup \{\rho\}$. By the definition of $\mathcal{I}_\mathcal{O}^r$, further, no element of $\mathsf{N}_\mathsf{I}^{\mathsf{aux}}$ can be an $\rho_\mathcal{O}$-successor. Hence, $r \in \mathsf{N}_\mathsf{R}$.

For (iii), by the same arguments as in (ii), we only have to consider the relation symbols in N_R and ρ. Elements of the form $x_{C,a}$ neither appear in the first position of a path, nor, by the definition of $\mathcal{I}_\mathcal{O}^r$, as r-successors, $r \in \mathsf{N}_\mathsf{R}$, and only have $\rho_\mathcal{O}$-predecessors of the form $x_{D,a} \in \mathsf{N}_\mathsf{I}^\rho$ or a. Hence, d_n must be of the form proposed if $\mathsf{Tail}(p) = x_{C,a}$. Item (iv) is analogous to (iii), and (v) holds by the definition of $\mathcal{I}_\mathcal{O}^r$. $\square$

The next Lemma 3.8 concretizes Proposition 3.7 even further, concerning the elements of $\Delta^{\mathcal{U}_\mathcal{O}}$ that belong to the equivalence class of some $p \in \Delta^{\mathcal{U}_\mathcal{O}}$ with $\mathsf{Tail}(p) \in \mathsf{N}_\mathsf{I}^{\mathsf{aux}}$. In particular, it restricts the kinds of paths occurring as elements of $\Delta^{\mathcal{U}_\mathcal{O}}$ that are indiscernible from p in $\mathcal{U}_\mathcal{O}$.

Lemma 3.8. *Let $p' \in [prx_C]_{\rho^{\mathcal{U}_\mathcal{O}}}$ with $x_C \in \mathsf{N}_\mathsf{I}^{\mathsf{aux}}$, and $r \in \mathsf{N}_\mathsf{R}$. Then either $p' = prx_C\rho\ell_{x_C}$ or p' is of the form $p' = prx_C(\rho x_{D_1,x_C}) \cdots (\rho x_{D_n,x_C})$ $n \geq 0$.*

Proof. By the definition of $\mathcal{U}_\mathcal{O}$, every $\rho_{\mathcal{O}'}$-successor of prx_C must be of the form $prx_C\rho d$, where d is either ℓ_{x_C} or of the form x_{D_0,x_C}. The latter have only $\rho_{\mathcal{O}'}$-successors of the form $prx_C\rho x_{D_0,x_C}\rho x_{D_1,x_C}$. $\square$

3.2.2 A mapping between the domain elements of $\mathcal{U}_\mathcal{O}$ and $\mathcal{I}_\mathcal{O}^r$

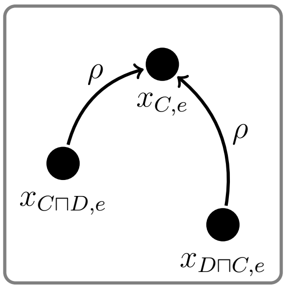
For each $p = d_0r_1d_1 \cdots r_nd_n \in \Delta^{\mathcal{U}_\mathcal{O}}$, we define a mapping $\mathsf{Tail}_{[p]} \colon [p]_{\rho^{\mathcal{U}_\mathcal{O}}} \to [d_n]_{\rho^{\mathcal{I}_\mathcal{O}^r}}$ given by $\mathsf{Tail}_{[p]}(q) = \mathsf{Tail}(q)$ for all $q \in [p]_{\rho^{\mathcal{U}_\mathcal{O}}}$. In what follows, we show that this function is well-defined and surjective.

Lemma 3.9. *For all $p, q \in \Delta^{\mathcal{U}_\mathcal{O}}$, if $(p, q) \in \rho^{\mathcal{U}_\mathcal{O}}$, then $(\mathsf{Tail}(p), \mathsf{Tail}(q)) \in \rho^{\mathcal{I}_\mathcal{O}^r}$.*

Proof. We prove this by induction on the construction of $\rho^{\mathcal{U}_\mathcal{O}}$. Assume first that $(p, q) \in \rho_{\mathcal{O}'}$. If $\mathsf{Tail}(p), \mathsf{Tail}(q) \in \mathsf{Ind}(\mathcal{A})$, we have $(\mathsf{Tail}(p), \mathsf{Tail}(q)) \in \rho^{\mathcal{I}_\mathcal{O}^r}$, by definition of $\mathcal{U}_\mathcal{O}$. Otherwise, we have $q = p \cdot \rho q'$, and the second line in the definition of $\rho_{\mathcal{O}'}$ implies $(\mathsf{Tail}(p), q') \in \rho_\mathcal{O}$. But then, $(\mathsf{Tail}(p), \mathsf{Tail}(q)) \in \rho^{\mathcal{I}_\mathcal{O}^r}$ since $\rho_\mathcal{O} \subseteq \rho^{\mathcal{I}_\mathcal{O}^r}$.

We now consider the induction steps of closing $\rho^{\mathcal{U}_\mathcal{O}}$ to an equivalence relation. Since $\rho^{\mathcal{I}_\mathcal{O}^r}$ is also an equivalence relation and, for any $d \in \Delta^{\mathcal{U}_\mathcal{O}}$, we obviously have $\mathsf{Tail}(d) \in \Delta^{\mathcal{I}_\mathcal{O}^r}$, reflexivity does not have to be considered further. For symmetry, we assume we have $(e, d) \in \rho^{\mathcal{U}_\mathcal{O}}$ and, by (IH), $(\mathsf{Tail}(e), \mathsf{Tail}(d)) \in \rho^{\mathcal{I}_\mathcal{O}^r}$. We then directly get $(\mathsf{Tail}(d), \mathsf{Tail}(e)) \in \rho^{\mathcal{I}_\mathcal{O}^r}$ since $\rho^{\mathcal{I}_\mathcal{O}^r}$ is an equivalence relation, either. The case for transitivity can be treated analogously. $\square$

The next lemma establishes surjectivity of Tail.



(b)

Figure 6: Example situation showing that Tail is not injective.

Lemma 3.10. *For all $p = d_0 r_1 d_1 \cdots r_n d_n \in \Delta^{\mathcal{U}_\mathcal{O}}$ and all $e' \in \Delta^{\mathcal{I}^r_\mathcal{O}}$, we have: If $(d_n, e') \in \rho^{\mathcal{I}^r_\mathcal{O}}$, then there is an element $q \in \Delta^{\mathcal{U}_\mathcal{O}}$ with $\mathsf{Tail}(q) = e'$ such that $(p, q) \in \rho^{\mathcal{U}_\mathcal{O}}$.*

Proof. For the proof we use induction on the construction of $\rho^{\mathcal{I}^r_\mathcal{O}}$. For the induction start, we assume $(d_n, e') \in \rho_\mathcal{O}$. By the definition of $\mathcal{I}^r_\mathcal{O}$, if $d_n, e' \in \mathsf{Ind}(\mathcal{A})$, then we have $(d_n, e') \in \rho_{\mathcal{O}'} \subseteq \rho^{\mathcal{U}_\mathcal{O}}$. Otherwise, if $e' \notin \mathsf{Ind}(\mathcal{A})$, since $p \in \Delta^{\mathcal{U}_\mathcal{O}}$ and $(d_n, e') \in \rho_\mathcal{O} \subseteq \rho^{\mathcal{I}^r_\mathcal{O}}$, we have the element $q = p \cdot \rho e' \in \Delta^{\mathcal{U}_\mathcal{O}}$ and hence get $(p, q) \in \rho_{\mathcal{O}'} \subseteq \rho^{\mathcal{U}_\mathcal{O}}$. We now consider the induction steps of closing $\rho^{\mathcal{I}^r_\mathcal{O}}$ to an equivalence relation. Since $\rho^{\mathcal{U}_\mathcal{O}}$ is also an equivalence relation, we obviously have $(d, d) \in \rho^{\mathcal{U}_\mathcal{O}}$ for all $(d_n, d_n) \in \rho^{\mathcal{I}^r_\mathcal{O}}$. For symmetry, we assume we have $(e', d_n) \in \rho^{\mathcal{I}^r_\mathcal{O}}$ and, by (IH), that for all $e \in \Delta^{\mathcal{U}_\mathcal{O}}$ with $\mathsf{Tail}(e) = e'$ there is some $d = d_0 r_1 d_1 \cdots r_n d_n \in \Delta^{\mathcal{U}_\mathcal{O}}$ such that $(e, d) \in \rho^{\mathcal{U}_\mathcal{O}}$. Since $\rho^{\mathcal{U}_\mathcal{O}}$ is also symmetric, this leads to $(d, e) \in \rho^{\mathcal{U}_\mathcal{O}}$. Again, the case for transitivity can be treated analogously. □

To note, however, that Tail' is not necessarily, consider Figure 6 as an extract of some canonical interpretation $\mathcal{I}_\mathcal{O}$ where both $x_{C\sqcap D,e}$ and $x_{D\sqcap C,e}$ are reachable. The two relations $\rho^{\mathcal{I}_\mathcal{O}}$ are clearly induced by the definition of $\mathcal{I}_\mathcal{O}$ and hence we also have the corresponding paths in $\mathcal{U}_\mathcal{O}$.

3.2.3 $\mathcal{U}_\mathcal{O}$ is suitable for query answering

We are now ready to show that the classical lemma about the connection between concept satisfiability in $\mathcal{U}_\mathcal{O}$ and $\mathcal{I}^r_\mathcal{O}$ also holds in our rough setting. This enables us to subsequently prove that $\mathcal{U}_\mathcal{O}$ is also a model of $\mathcal{O}$ and especially suitable for query answering.

Lemma 3.11. *For all concepts C and $p = d_0 r_1 d_1 \cdots r_n d_n \in \Delta^{\mathcal{U}_\mathcal{O}}$, $p \in C^{\mathcal{U}_\mathcal{O}}$ iff $d_n \in C^{\mathcal{I}^r_\mathcal{O}}$.*

Proof. The proof is by induction on the structure of C. If $C = A \in \mathsf{N_C}$, the claim directly follows, by the definition of $\mathcal{U}_\mathcal{O}$. The case $C = C_1 \sqcap C_2$ follows easily by definition. We now consider the case $C = \exists r.C_1$.

($\Rightarrow$) If $p \in (\exists r.C_1)^{\mathcal{U}_\mathcal{O}}$, then there is a $q \in \Delta^{\mathcal{U}_\mathcal{O}}$ such that $(p, q) \in r^{\mathcal{U}_\mathcal{O}}$ and $q \in C_1^{\mathcal{U}_\mathcal{O}}$. By definition of $\mathcal{U}_\mathcal{O}$, either (i) $p, q \in \mathsf{Ind}(\mathcal{A})$ and $(p, q) \in r^{\mathcal{I}^r_\mathcal{O}}$, or (ii) q is of the form $q = p{\cdot}se$ for some role s with $\mathcal{R} \models s \sqsubseteq r$. For the latter, we get $p{\cdot}se \in \mathsf{Paths}_\mathcal{A}(\mathcal{I}^r_\mathcal{O})$, by the definition of $\mathcal{U}_\mathcal{O}$, which leads to $(d_n, e) \in s^{\mathcal{I}^r_\mathcal{O}}$. By induction, $\mathsf{Tail}(q) \in C_1^{\mathcal{I}^r_\mathcal{O}}$, in both cases, and $d_n \in (\exists r.C_1)^{\mathcal{I}^r_\mathcal{O}}$ follows.

($\Leftarrow$) If $d_n \in (\exists r.C_1)^{\mathcal{I}^r_\mathcal{O}}$, then there is an $e' \in C_1^{\mathcal{I}^r_\mathcal{O}}$ with $(d_n, e) \in r^{\mathcal{I}^r_\mathcal{O}}$. By definition of $\mathcal{I}^r_\mathcal{O}$, either $e \in \mathsf{Ind}(\mathcal{A}) \subseteq \Delta^{\mathcal{U}_\mathcal{O}}$, or there is a role s and a $q \in \Delta^{\mathcal{U}_\mathcal{O}}$ such that $\mathcal{R} \models s \sqsubseteq r$ and, $q = p \cdot se'$ hold. For both cases, the definition of $\mathcal{U}_\mathcal{O}$ yields $(p, q) \in r^{\mathcal{U}_\mathcal{O}}$. By induction, $q \in C_1^{\mathcal{U}_\mathcal{O}}$, and $p \in (\exists r.C_1)^{\mathcal{U}_\mathcal{O}}$ follows.

We now focus on the rough constructors. Consider first $C = \overline{C_1}$.

($\Rightarrow$) Let $p \in \overline{C_1}^{\mathcal{U}_\mathcal{O}}$, then there is some $q \in \Delta^{\mathcal{U}_\mathcal{O}}$ such that $(p, q) \in \rho^{\mathcal{U}_\mathcal{O}}$ and $q \in C_1^{\mathcal{U}_\mathcal{O}}$. By Lemma 3.9, we get $(d_n, \mathsf{Tail}(q)) \in \rho^{\mathcal{I}^r_\mathcal{O}}$, and by induction $\mathsf{Tail}(e) \in C_1^{\mathcal{I}^r_\mathcal{O}}$. Hence, $d_n \in \overline{C_1}^{\mathcal{I}^r_\mathcal{O}}$.

($\Leftarrow$) If $d_n \in \overline{C_1}^{\mathcal{I}^r_\mathcal{O}}$, then there is some $e' \in C_1^{\mathcal{I}^r_\mathcal{O}}$ with $(d_n, e') \in \rho^{\mathcal{I}^r_\mathcal{O}}$. By Lemma 3.10, there is a $q \in \Delta^{\mathcal{U}_\mathcal{O}}$ such that $(p, q) \in \rho^{\mathcal{U}_\mathcal{O}}$ and $\mathsf{Tail}(q) = e'$. By induction, $q \in C_1^{\mathcal{U}_\mathcal{O}}$, and thus $p \in \overline{C_1}^{\mathcal{U}_\mathcal{O}}$.

The only remaining case is $C = \underline{C_1}$.

($\Rightarrow$) If $p \in \underline{C_1}^{\mathcal{U}_\mathcal{O}}$, then for all $q \in [p]_{\rho^{\mathcal{U}_\mathcal{O}}}$ we have $q \in C_1^{\mathcal{U}_\mathcal{O}}$ and by induction $\mathsf{Tail}(q) \in C_1^{\mathcal{I}^r_\mathcal{O}}$. By Lemma 3.10, for every $e \in [d_n]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$, there is a $q' \in [p]_{\rho^{\mathcal{U}_\mathcal{O}}}$ such that $\mathsf{Tail}(q') = e$, and hence $e \in C_1^{\mathcal{I}^r_\mathcal{O}}$. This implies $d_n \in \underline{C_1}^{\mathcal{U}_\mathcal{O}}$.

($\Leftarrow$) If $d_n \in \underline{C_1}^{\mathcal{I}^r_\mathcal{O}}$, then for all $e' \in [d_n]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$ we have $e' \in C_1^{\mathcal{I}^r_\mathcal{O}}$. Let $q \in [p]_{\rho^{\mathcal{U}_\mathcal{O}}}$. By Lemma 3.9 we know that $\mathsf{Tail}(q) \in [d_n]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$. By induction, $q \in C_1^{\mathcal{U}_\mathcal{O}}$, and hence $p \in \underline{C_1}^{\mathcal{U}_\mathcal{O}}$. □

Finally, using Lemmas 3.5 and 3.11, it is straightforward to establish the following.

Lemma 3.12. *If $\mathcal{O}$ is consistent, then*

1. *$\mathcal{U}_\mathcal{O}$ is a model of $\mathcal{O}$, and*
2. *for all k-ary CQs ψ and $a_1, \ldots, a_k \in \mathsf{Ind}(\mathcal{O})$, $(a_1, \ldots, a_k) \in \mathsf{Cert}(\psi, \mathcal{O})$ iff $\mathcal{U}_\mathcal{O} \models \psi(a_1, \ldots, a_k)$.*

Proof. The first point is a direct consequence of Lemmas 3.5 and 3.11. For the second point, the *only if* direction follows trivially from the first point. We now prove the *if* direction.

Assume that $\mathcal{U}_\mathcal{O} \models \psi(a_1, \ldots, a_k)$ holds and let $\mathcal{I}$ be an arbitrary model of $\mathcal{O}$. We define a mapping $\pi : \Delta^{\mathcal{U}_\mathcal{O}} \to \Delta^\mathcal{I}$ such that for all $p, p' \in \Delta^{\mathcal{U}_\mathcal{O}}$, $a \in \mathsf{Ind}(\mathcal{A})$, roles r and concepts C:

1. $\pi(a) = a^\mathcal{I}$;
2. $p \in C^{\mathcal{U}_\mathcal{O}}$ implies $\pi(p) \in C^\mathcal{I}$;
3. $(p, p') \in r^{\mathcal{U}_\mathcal{O}}$ implies $(\pi(p), \pi(p')) \in r^\mathcal{I}$; and
4. $(p, p') \in \rho^{\mathcal{U}_\mathcal{O}}$ implies $(\pi(p), \pi(p')) \in \rho^\mathcal{I}$.

We define π by induction on the structure of the path. If p is of the form a for some $a \in \mathsf{Ind}(\mathcal{A})$, then set $\pi(a) := a^\mathcal{I}$ as required by the first condition. It is easy to see that all other conditions are also satisfied.

Let now $p = p'rd \in \Delta^{\mathcal{U}_\mathcal{O}}$, where $r \in \mathsf{N_R} \cup \{\rho\}$, and let $d' = \mathsf{Tail}(p')$. If $r \in \mathsf{N_R}$, then $(d', d) \in r^{\mathcal{I}^r_\mathcal{O}}$ and d is of the form $x_D \in \mathsf{N_I^{aux}}$. By Lemma 3.5, $x_D \in D^{\mathcal{I}^r_\mathcal{O}}$, and hence $d' \in (\exists r.D)^{\mathcal{I}^r_\mathcal{O}}$. By Lemma 3.11 it follows that $p' \in (\exists r.D)^{\mathcal{U}_\mathcal{O}}$ and by induction $\pi(p') \in (\exists r.D)^\mathcal{I}$. Hence there exists an $e \in \Delta^\mathcal{I}$ with $(\pi(p'), e) \in r^\mathcal{I}$ and $e \in D^\mathcal{I}$. We define $\pi(p) := e$. Alternatively, if $r = \rho$, then $(d', d) \in \rho_\mathcal{O}$, and by Proposition 3.7, $d \in \mathsf{N_I^\rho}$. If d is of the form $x_{D,y} \in \mathsf{N_I^{up}}$, then $d' \in \overline{D}^{\mathcal{I}^r_\mathcal{O}}$. As before, this implies that $\pi(p') \in \overline{D}^\mathcal{I}$, and hence, there exists an $e \in \Delta^\mathcal{I}$ such that $(\pi(p'), e) \in \rho^\mathcal{I}$ and $e \in D^\mathcal{I}$. We define $\pi(p) := e$. Finally, if d is of the form $\ell_y \in \mathsf{N_I^{low}}$, then we set $\pi(p) := \pi(p')$.

As a result, π satisfies Properties 1. and 3. by construction. For 4. this is the case because $\mathcal{I}$ also must interpret ρ as equivalence relation. To see that 2. is also satisfied, let $p = d_0r_1d_1 \cdots r_nd_n \in C^{\mathcal{U}_\mathcal{O}}$. If $d_n = x_{D,x_E}$ for some $x_E \in \mathsf{N_I^{aux}}$, then by Lemmas 3.5 and 3.11, either (i) $\mathcal{O} \models D \sqsubseteq C$ or (ii) $\mathcal{O} \models E \sqsubseteq \underline{C}$. By construction, we have $\pi(x_{D,x_E}) \in D^\mathcal{I}$ and hence get $\pi(x_{D,x_E}) \in C^\mathcal{I}$ for (i), by the semantics. For (ii), we have that d is of the form $d = p \cdot x_E(\rho x_{D'_i,x_E})^i \rho x_{D,x_E}$, by Proposition 3.7, and also $p \cdot x_E \rho^{\mathcal{U}_\mathcal{O}} d$, by the definition of $\mathcal{U}_\mathcal{O}$. Further, Lemmas 3.5 and 3.11 imply $p \cdot x_E \in E^{\mathcal{U}_\mathcal{O}}$. We thus have some element $e \in \Delta^\mathcal{I}$ with $e \in E^\mathcal{I}$, by (IH 2), for which we have $\pi(d)\rho^\mathcal{I}e$, by (IH 4). But then, we get $\pi(d) \in C^\mathcal{I}$ for (i), by the semantics, too. The case for $d_n = x_{D,a}, a \in \mathsf{Ind}(\mathcal{A})$ can be shown correspondingly. The case for $d_n \in \mathsf{N_I^{aux}}$ and $d_n \in \mathsf{N_I^{low}}$ is similar to (i) and (ii), respectively. □

This section covered the first extension of the combined approach of query answering [LTW09] to the rough setting, namely that of the definition of the canonical interpretation together with that of the unraveling. Due to the special semantics of the indiscernibility relation, we had to take specifically care that our extensions of the classical constructions also retained their properties –because the latter are integral to prove the correctness of the query rewriting. Our extension of that rewriting is described in the next section.

4 Answering Rough Queries

In the previous section we showed that all rough conjunctive queries can be answered over the unraveled canonical model $\mathcal{U}_\mathcal{O}$. However, since $\mathcal{U}_\mathcal{O}$ is infinite, it cannot be used as a tool for effective query answering. It was shown in [LTW09] that it is possible to rewrite conjunctive queries written in $\mathcal{ELH}_\bot$ in such a way that they can be answered over the polynomially large canonical interpretation directly. The main idea of this combined approach is to simulate the unraveling of this model by disallowing merging paths over the anonymous variables. We show that this idea can be extended to the rough DL $\mathcal{ELH}_{\bot\rho}$. Our goal is to prove the following theorem.

Theorem 4.1. *For every finite set of role inclusions $\mathcal{R}$ and k-ary CQ ϕ, it is possible to construct in polynomial time a k-ary FO query $\phi^\dagger_\mathcal{R}$ such that for all $\mathcal{ELH}_{\bot\rho}$-ontologies $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{R}$ the set of role inclusions in $\mathcal{T}$ and all $a_1, \dots, a_k \in \mathsf{Ind}(\mathcal{A})$, we have $(a_1, \dots, a_k) \in \mathsf{Cert}(\phi, \mathcal{O})$ iff $\mathcal{I}^r_\mathcal{O} \models \phi^\dagger_\mathcal{R}(a_1, \dots, a_k)$.*

This theorem states that it is possible to transform (or *rewrite*) any CQ ϕ into a FO-query $\phi^\dagger_\mathcal{R}$ that can be answered over the canonical model $\mathcal{I}^r_\mathcal{O}$. Recall from Definition 2.5 that FO queries generalize CQs. The rest of this section is dedicated to proving it.

Let now $\mathcal{R}$ and ϕ be an arbitrary but fixed finite set of role inclusions, and a k-ary CQ, respectively. We construct the FO-query $\phi^\dagger_\mathcal{R}$. To do this, we first introduce two additional unary predicates (i.e., concepts) Aux and Aux_ρ, and one new binary predicate ρ_L which we assume to be always interpreted by the canonical interpretation $\mathcal{I}^r_\mathcal{O}$ and its unraveling $\mathcal{U}_\mathcal{O}$, as follows:

$$\begin{aligned}
\mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}} &:= \Delta^{\mathcal{I}^r_\mathcal{O}} \cap \mathsf{N}_\mathsf{I}^{\mathsf{aux}} \\
\mathsf{Aux}_\rho^{\mathcal{I}^r_\mathcal{O}} &:= \Delta^{\mathcal{I}^r_\mathcal{O}} \cap \mathsf{N}_\mathsf{I}^{\rho} \\
\rho_L^{\mathcal{I}^r_\mathcal{O}} &:= \{(e, \ell_e) \in \Delta^{\mathcal{I}^r_\mathcal{O}} \times \mathsf{N}_\mathsf{I}^{\mathsf{low}}\} \\
\mathsf{Aux}^{\mathcal{U}_\mathcal{O}} &:= \{p \in \Delta^{\mathcal{U}_\mathcal{O}} \mid \mathsf{Tail}(p) \in \mathsf{N}_\mathsf{I}^{\mathsf{aux}}\} \\
\mathsf{Aux}_\rho^{\mathcal{U}_\mathcal{O}} &:= \{p \in \Delta^{\mathcal{U}_\mathcal{O}} \mid \mathsf{Tail}(p) \in \mathsf{N}_\mathsf{I}^{\rho}\} \\
\rho_L^{\mathcal{U}_\mathcal{O}} &:= \{(p \cdot e, p \cdot e\rho\ell_e) \in \Delta^{\mathcal{U}_\mathcal{O}} \times \Delta^{\mathcal{U}_\mathcal{O}}\}.
\end{aligned}$$

These predicates are *new* in the sense that they occur neither in $\mathcal{A}$ nor in ϕ.

Let $\sim_\phi$ denote the equivalence relation over $\mathsf{Term}(\phi)$ induced by the atoms of the form $\rho(s,t)$ occurring in ϕ, and $\sim_{\phi^r}$ denote the smallest transitive and reflexive relation on $\mathsf{Term}(\phi)$ that includes the relation:

$$\{(t, t') \mid r_1(s,t), r_2(s',t') \in \phi, r_1, r_2 \in \mathsf{N}_\mathsf{R}, t \sim_\phi t'\},$$

and satisfies the closure condition:

$$\text{if } r_1(s,t), r_2(s',t') \in \phi \text{ and } t \sim_{\phi^r} t', \text{ then } s \sim_{\phi^r} s'. \qquad (*)$$

The relation $\sim_{\phi^r}$ will be fundamental for constructing $\phi^{\dagger}_{\mathcal{R}}$. Intuitively, the equivalence classes of $\sim_{\phi^r}$ contain those terms that are not discerned by any match for $\mathcal{U}_{\mathcal{O}}$ and ϕ; i.e., they are always mapped to the same element in $\Delta^{\mathcal{U}_{\mathcal{O}}}$. This is obviously the case for identical terms. To understand the closure condition $(*)$, first assume that $t = t'$; then $(*)$ describes a non-tree situation in the query ϕ since the term t has two predecessors s and s'. Therefore, any match of the query in $\mathcal{I}^r_{\mathcal{O}}$ that maps t to the Aux-part should map s and s' to the same element; otherwise this match would not exist in the unraveled model $\mathcal{U}_{\mathcal{O}}$. The case where $t \sim_{\phi^r} t'$ can be understood analogously. Since any match for $\mathcal{I}^r_{\mathcal{O}}$ and ϕ maps t and t' to the same element if $t \sim_{\phi} t'$ and t is mapped to the Aux-part (see Lemma 3.4) we include such t, t' into $\sim_{\phi^r}$.

For any equivalence class ζ of $\sim_{\phi^r}$, we define the sets:

$$\mathsf{Pre}(\zeta) := \{t \mid r(t,t') \in \phi, r \in \mathsf{N_R}, t' \in \zeta\}, \text{ and}$$
$$\mathsf{In}(\zeta) := \{r \mid r(t,t') \in \phi, t \in \mathsf{Term}(\phi), t' \in \zeta\}$$

The role $r \in \mathsf{N_R}$ is called an *implicant* of $R \subseteq \mathsf{N_R}$ if $\mathcal{R} \models r \sqsubseteq s$ for all $s \in R$. It is called a *prime implicant* if, additionally, $\mathcal{R} \not\models r \sqsubseteq r'$ for all implicants r' of R with $r \neq r'$. By the assumption 2 made in Section 2 (see Page 7), there is a prime implicant for any set $R \in \mathsf{N_R}$ for which there is an implicant. We define the following auxiliary sets:

- $\mathsf{Fork}_{=}$ is the set of pairs $(\mathsf{Pre}(\zeta), \zeta)$ with $\mathsf{Pre}(\zeta)$ of cardinality at least two;
- $\mathsf{Fork}_{\neq}$ is the set of variables $v \in \mathsf{QVar}(\phi)$ such that there is no implicant of $\mathsf{In}([v]_{\sim_{\phi^r}})$;
- $\mathsf{Fork}_{\mathcal{H}}$ is the set of pairs (I, ζ) such that $\mathsf{Pre}(\zeta) \neq \emptyset$, there is a prime implicant of $\mathsf{In}(\zeta)$ that is not contained in $\mathsf{In}(\zeta)$, and I is the set of all prime implicants of $\mathsf{In}(\zeta)$;
- Cyc is the set of all variables $v \in \mathsf{QVar}(\phi)$ such that there exist atoms $r_0(t_0, t'_0), \ldots, r_m(t_m, t'_m)$, $m \geq 0$ in ϕ with $r_i \in \mathsf{N_R}$, and (i) $(v, t_i) \in {\sim_{\phi^r}} \cup {\sim_{\phi}}$ for some $i \leq m$, (ii) $(t'_i, t_{i+1}) \in {\sim_{\phi^r}} \cup {\sim_{\phi}}$ for all $i < m$, and (iii) $t'_m \sim_{\phi} t_0$.

For each equivalence class ζ of $\sim_{\phi^r}$, we select an arbitrary but fixed representative $t_\zeta \in \zeta$, and if $\mathsf{Pre}(\zeta) \neq \emptyset$, we also select a fixed $t^{\mathsf{Pre}}_{\zeta} \in \mathsf{Pre}(\zeta)$.

Before continuing with the proof, we provide the basic motivation behind these definitions. Intuitively, the first element of a pair in $\mathsf{Fork}_{=}$ describes the variables that are mapped to the same element by any match for ϕ and $\mathcal{U}_{\mathcal{O}}$, which maps the successor variable to the Aux-part. The variables which, by the design of ϕ (i.e., the role atoms where the variables occur as successors) and the construction of $\mathcal{U}_{\mathcal{O}}$, can never be mapped to such a common successor are collected in $\mathsf{Fork}_{\neq}$.

$$
\begin{array}{llcl}
(UF1) & \overline{C}(x) & \rightarrow & \exists y.\rho(x,y) \land C(y) \\
(UF2) & \underline{C}(x) & \rightarrow & \exists y_1, y_2.\rho(x,y_1) \land \rho_L(y_1,y_2) \land C(y_2) \\
(UF3) & C \sqcap D(x) & \rightarrow & C(x) \land D(x) \\
(UF4) & \exists r.C(x) & \rightarrow & \exists y.r(x,y) \land C(y), r \in \mathsf{N_R}
\end{array}
$$

Figure 7: Unfolding rules for constructing ψ_{CQ}

Let now $\phi = \exists \vec{x}.\psi$ be a CQ. We construct the FOL query ψ_{CQ} by exhaustively applying the unfolding rules in Figure 7, where a rule applcation corresponds to replacing a conjunction on the left-hand side of the rule, by those on the right-hand side. In the rules, C and D denote arbitrary complex concepts, and y_1, y_2, and y fresh variables for each rule application.

The FOL rewriting $\phi^{\dagger}_{\mathcal{R}}$ of ϕ is defined as $\exists \vec{x}.(\psi_{\mathsf{CQ}} \land \psi_0 \land \psi_1 \land \psi_2 \land \psi_3)$, where ψ_0, ψ_1, ψ_2, and ψ_3 are given by

$$
\begin{aligned}
\psi_0 &:= \bigwedge_{v \in \mathsf{AVar}} \neg \mathsf{Aux}_{\rho}(v) \\
\psi_1 &:= \bigwedge_{v \in \mathsf{AVar} \cup \mathsf{Fork}_{\neq} \cup \mathsf{Cyc}} \neg \mathsf{Aux}(v) \\
\psi_2 &:= \bigwedge_{(\{t_1, \dots t_k\}, \zeta) \in \mathsf{Fork}_{=}} (\mathsf{Aux}(t_\zeta) \rightarrow \bigwedge_{1 \leq i < k} t_i = t_{i+1}) \\
\psi_3 &:= \bigwedge_{(\mathsf{I}, \zeta) \in \mathsf{Fork}_{\mathcal{H}}} (\mathsf{Aux}(t_\zeta) \rightarrow \bigvee_{r \in \mathsf{I}} r(t_\zeta^{\mathsf{Pre}}, t_\zeta))
\end{aligned}
$$

Notice that the terms used in this construction are based on the original query ϕ, and hence do not apply to the existentially quantified variables introduced during the application of the unfolding rules in the construction of ψ_{CQ}.

To help understanding this rewriting procedure, we now provide some simple examples of their application. The first example demostrates the role of the set Cyc in the rewriting, while Example 4.3 shows the use of $\mathsf{Fork}_{=}$. For these examples, we consider $\mathcal{R} = \emptyset$. Notice that in this case, $\mathsf{Fork}_{\mathcal{H}}$ is always empty, and hence $\psi_3 = \mathsf{true}$. Thus, we ommit this formulas in them.

Example 4.2. *Let*

$$\phi = \exists y_1, y_2.(hasA(y_1, y_2) \land \rho(y_1, y_2)).$$

We have $\mathsf{Cyc} = \{y_1, y_2\}$, $\mathsf{Fork}_{=} = \mathsf{Fork}_{\neq} = \mathsf{Fork}_{\mathcal{H}} = \emptyset$, *and thus obtain*

$$\phi^{\dagger}_{\mathcal{R}} = \exists y_1, y_2.(hasA(y_1, y_2) \land \rho(y_1, y_2) \land \neg \mathsf{Aux}(y_1) \land \neg \mathsf{Aux}(x_2)).$$

This query guarantees that all the answer pairs provided are indiscernible elements, related via the role hasA, but additionally, none of the auxiliary elements produced in the construction of the canonical interpretation is returned.

For the next example, we have a similar query, demostrating the rewriting of forking situations.

Example 4.3. *Let*

$$\phi(x_1, x_2) = \exists y_1, y_2.(hasA(x_1, y_1) \wedge hasA(x_2, y_2) \wedge \rho(y_1, y_2))$$

The relation $\sim_\phi$ *has the equivalence classes* $\{x_1\}, \{x_2\}$, *and* $\{y_1, y_2\}$, *and* $\sim_{\phi^r}$ *has the equivalence classes* $\{x_1, x_2\}$ *and* $\{y_1, y_2\}$. *Moreover,* $\mathsf{Pre}(\{y_1, y_2\}) = \{x_1, x_2\}$ *and* $\mathsf{In}(\{y_1, y_2\}) = \{hasA\}$. *Thus, we have that* $\mathsf{Fork}_= = \{(\{x_1, x_2\}, \{y_1, y_2\})\}$, *and* $\mathsf{Fork}_{\neq} = \mathsf{Fork}_{\mathcal{H}} = \mathsf{Cyc} = \emptyset$. *With this, we obtain the rewriting*

$$\begin{aligned}\phi^\dagger_{\mathcal{R}} = \exists y_1, y_2.(\quad & hasA(x_1, y_1) \wedge hasA(x_2, y_2) \wedge \rho(y_1, y_2) \wedge \\ & \neg\mathsf{Aux}(x_1) \wedge \neg\mathsf{Aux}(x_1) \wedge (\mathsf{Aux}(y_1) \rightarrow x_1 = x_2)).\end{aligned}$$

The last conjunct avoids the joining of two different individuals of the ABox through auxiliar elements of the canonical interpretation.

Let now $\phi = \exists \vec{x}.\psi$ be a CQ, and π a valuation of the variables in ϕ such that $\mathcal{U}_\mathcal{O} \models^\pi \psi$. We define the mapping $\tau : \mathsf{Term}(\phi^\dagger_{\mathcal{R}}) \rightarrow \Delta^{\mathcal{I}^r_\mathcal{O}}$ inductively on the application of the unfolding rules from Figure 7 as follows:

- $\tau(t) = \mathsf{Tail}(\pi(t))$ for all $t \in \mathsf{Term}(\phi)$;
- if $\rho(x, y) \wedge C(y)$ was introduced by (UF1), then $\tau(y) = x_{C,b}$ if $\tau(x)$ is of the form $b, x_{D,b}$, or ℓ_b, $b \in \mathsf{Ind}(\mathcal{A})$, and $\tau(y) = x_{C,x_D}$ if $\tau(x) \in [x_D]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$;
- if $\rho(x, y_1) \wedge \rho_L(y_1, y_2)$ was introduced by (UF2) then
  - $\tau(y_1) = b, \tau(y_2) = \ell_b$ if $\tau(x)$ is of the form $b, x_{D,b}$, or ℓ_b, $b \in \mathsf{Ind}(\mathcal{A})$, and
  - $\tau(y_1) = x_C, \tau(y_2) = \ell_{x_C}$ if $\tau(x) \in [x_D]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$; and
- if $r(x, y) \wedge C(y)$ was introduced by (UF4), then $\tau(y) = x_C$

It is easy to see that this function τ is well defined. We now show that $\mathcal{I}^r_\mathcal{O} \models^\tau \psi_{\mathsf{CQ}}$.

Lemma 4.4. *Let* $\phi = \exists \vec{x}.\psi$ *be a CQ, and* $\mathcal{U}_\mathcal{O} \models^\pi \psi$. *Then* $\mathcal{I}^r_\mathcal{O} \models^\tau \psi_{\mathsf{CQ}}$.

Proof. We prove this by induction on the application of unfolding rules for constructing ψ_{CQ}. Let $\psi^0, \psi^1, \ldots$ be the sequence queries obtained at each application of an unfolding rule, with $\psi^0 = \psi$. For the best case, it follows directly from Lemma 3.11 and the construction of τ that $\mathcal{I}^r_\mathcal{O} \models \tau\psi = \psi^0$. Suppose now that $\mathcal{I}^r_\mathcal{O} \models^\tau \psi^n$, we prove that $\mathcal{I}^r_\mathcal{O} \models^\tau \psi^{n+1}$ by a case analysis over the rule applied.

(UF1) ψ^{n+1} is obtained from ψ^n by replacing $\overline{C}(x)$ by $\exists y.\rho(x, y) \wedge C(y)$, where $x \in \mathsf{Term}(\psi^n)$. By induction, we know that $\tau(x) \in \exists\rho.C^{\mathcal{I}^r_\mathcal{O}}$. Lemma 3.4, $\tau(x)$

can only be an equivalence class of the form $[b]_{\rho^{\mathcal{I}^r_{\mathcal{O}}}}$, $b \in \mathsf{Ind}(\mathcal{A})$, or $[x_D]_{\rho^{\mathcal{I}^r_{\mathcal{O}}}}$, $x_D \in \mathsf{N}_\mathsf{I}^{\mathsf{aux}}$. From Lemma 3.5 it then follows that $\mathcal{O} \models \exists\rho.C(a)$ or $\mathcal{O} \models D \sqsubseteq \exists\rho.C$, respectively. But then $(\tau(x), x_{C,e}) \in \rho^{\mathcal{I}^r_{\mathcal{O}}}$ and $x_{C,e} \in C^{\mathcal{I}^r_{\mathcal{O}}}$, where e is either b or x_D, respectively. This implies that $\mathcal{I}^r_{\mathcal{O}} \models^\tau \psi^{n+1}$.

The other cases can be shown similarly. $\square$

This lemma shows that τ is an $(a_1,\ldots,a_k)$-match for $\mathcal{I}^r_{\mathcal{O}}$ and ψ_{CQ}. Since our goal is to show that it is a match for $\phi^\dagger_{\mathcal{R}}$, we need to prove that $\mathcal{I}^r_{\mathcal{O}} \models^\tau \psi_i$ for all $i, 0 \leq i \leq 3$. Notice that all the new variables introduced to $\phi^\dagger_{\mathcal{R}}$ during the rewriting are existentially quantified, and hence cannot be answer variables; moreover, the auxiliary sets $\mathsf{Fork}_=, \mathsf{Fork}_{\neq}, \mathsf{Fork}_{\mathcal{H}}$, and Cyc used are defined w.r.t. the relation $\sim_{\phi^r}$. Thus, it suffices to consider only $\tau(t)$ for $t \in \mathsf{Term}(\phi)$. We start by showing the following result.

Lemma 4.5. *Let $s, t \in \mathsf{Term}(\phi)$ be such that $s \sim_{\phi^r} t$ and $\pi(s) \in \mathsf{Aux}^{\mathcal{U}_{\mathcal{O}}}$. Then*

1. *$\pi(s) = \pi(t)$ and*
2. *for all terms s', t' and roles r_1, r_2, if $r_1(s', s), r_2(t', t) \in \phi$, then $\pi(s') = \pi(t')$.*

Proof. By definition, $\sim_{\phi^r}$ is the smallest transitive and reflexive relation that includes $\{(t,t') \mid r_1(s,t), r_2(s',t') \in \phi, r_1, r_2 \in \mathsf{N}_\mathsf{R}, t \sim_\phi t'\}$, and is closed under $(*)$ (see page 22).

We prove 1 by induction on the definition of $\sim_{\phi^r}$. If $s \sim_{\phi^r} t$, $s \neq t$ then there exist $r_1(s', s), r_2(t', t)$ such that $s \sim_\phi t$. Since π is a match for ϕ and $\mathcal{U}_{\mathcal{O}}$, we have that $\pi(s), \pi(t) \in \mathsf{Ind}(\mathcal{A}) \cup \mathsf{Aux}^{\mathcal{U}_{\mathcal{O}}}$. By assumption, $\pi(s) \in \mathsf{Aux}^{\mathcal{U}_{\mathcal{O}}}$, and from Lemma 3.4 we get $\pi(s) = \pi(t)$. The result follows trivially for the reflexive closure. We only need to prove it for the closure under transitivity and $(*)$.

Assume that the result holds for $s \sim_{\phi^r} t'$ and $t' \sim_{\phi^r} t$. Then, by induction hypothesis, $\pi(s) = \pi(t') = \pi(t)$. Suppose now that $r_1(s, s'), r_2(t, t') \in \phi$ and the result holds for $s' \sim_{\phi^r} t'$. Since $(\pi(s), \pi(s')) \in r_1^{\mathcal{U}_{\mathcal{O}}}$, $\pi(s') \in \mathsf{Aux}^{\mathcal{U}_{\mathcal{O}}}$, and hence, by induction, $\pi(s') = \pi(t')$. But then, by the construction of the unraveled interpretation, $\pi(s) = \pi(t)$.

The property 2 follows directly from 1 and the closure under $(*)$. $\square$

Using this result, we can then show that τ is a match for the auxiliary queries ψ_i.

Lemma 4.6. *If $\mathcal{U}_{\mathcal{O}} \models^\pi \psi$, then $\mathcal{I}^r_{\mathcal{O}} \models^\tau \psi_i$ for $i, 0 \leq i \leq 3$.*

Proof. For $\psi_0 = \bigwedge_{v \in \mathsf{AVar}} \neg\mathsf{Aux}_\rho(v)$, let $v \in \mathsf{AVar}$. Then, by definition of query answers, $\pi(v) \in \mathsf{Ind}(\mathcal{A})^{\mathcal{U}_{\mathcal{O}}}$. But then, $\tau(v) = \pi(v) \in \mathsf{Ind}(\mathcal{A})^{\mathcal{U}_{\mathcal{O}}} = \mathsf{Ind}(\mathcal{A})^{\mathcal{I}^r_{\mathcal{O}}}$, and hence $\tau(v) \notin \mathsf{Aux}_\rho^{\mathcal{I}^r_{\mathcal{O}}}$.

We now consider the case of ψ_1. If $v \in \mathsf{AVar}(\phi)$, then as in the previous case, $\tau(v) \notin \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}}$. Suppose now that $\tau(v) \in \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}}$. If $v \in \mathsf{Fork}_{\neq}$ then, there is no implicant for $\mathsf{In}([v]_{\sim_{\phi^r}})$. By definition, for every $r \in \mathsf{In}([v]_{\sim_{\phi^r}})$ there exists $r(s_r, t_r) \in \phi$ such that $t_r \sim_{\phi^r} v$. Moreover, since $\tau(v) \in \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}}$, Lemma 3.11 implies that $\pi(v) \in \mathsf{Aux}^{\mathcal{U}_\mathcal{O}}$; thus $\pi(v) = \pi(t_r)$ (Lemma 4.5), and $(\pi(s_r), \pi(v)) \in r^{\mathcal{U}_\mathcal{O}}$. By the unraveling condition, this implies that for all $r, r' \in \mathsf{In}([v]_{\sim_{\phi^r}})$ $\pi(s_r) = \pi(s_{r'})$; but then every $r \in \mathsf{In}([v]_{\sim_{\phi^r}})$ is an implicant for $\mathsf{In}([v]_{\sim_{\phi^r}})$, yielding a contradiction. Finally, if $v \in \mathsf{Cyc}$ then there exist $r_i(t_i, t'_i) \in \phi, 0 \leq i \leq m$ and $j, 0 \leq j \leq m$ with $(v, t_j) \in \sim_{\phi^r} \cup \sim_\phi$. Since $\tau(v) \in \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}}$, it follows from Lemmas 4.5 and 3.8 that $\pi(t_j) \in \mathsf{Aux}^{\mathcal{U}_\mathcal{O}}$, and therefore $\pi(t'_j) = \pi(t_j) \cdot rd$ for some $d \in \Delta^{\mathcal{I}^r_\mathcal{O}}$. In particular, $\pi(t'_j) \in \mathsf{Aux}^{\mathcal{U}_\mathcal{O}}$. Additionally, we know that $(t'_i, t_{i+1}) \in \sim_{\phi^r} \cup \sim_\phi$ for all $i, 0 \leq i < m$ and $(t'_m, t_0) \in \sim_{\phi^r} \cup \sim_\phi$. Repeating the previous argument, we obtain that $\pi(t_j) = \pi(t_{j+m \mod m+1}) = \pi(t_j) r_j p$ for some path p, which is a contradiction.

To prove that it is a match of ψ_2, let $(\{t_1, \ldots, t_k\}, \zeta) \in \mathsf{Fork}_=$ such that $t_\zeta \in \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}}$. Then, $\pi(t_\zeta) \in \mathsf{Aux}^{\mathcal{U}_\mathcal{O}}$ and there are terms $t'_1, \ldots, t'_k \in \zeta$ and role names $r_1, \ldots, r_k$ such that $r_i(t_i, t'_i) \in \phi$ for all $i, 1 \leq i \leq k$. By Lemma 4.5 (2) $\pi(t_i) = \pi(t_j)$ and hence $\tau(t_i) = \tau(i_j)$ holds for all $1 \leq i, j \leq k$.

Finally, we prove the claim for ψ_3. Let $(\mathsf{I}, \zeta) \in \mathsf{Fork}_\mathcal{H}$ such that $\tau(t_\zeta) \in \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}}$. Since $\mathsf{Pre}(\zeta) \neq \emptyset$, t^{Pre}_ζ is defined and $\Gamma := \{r \in \mathsf{N_R} \mid (\tau(t^{\mathsf{Pre}}_\zeta), \tau(t_\zeta)) \in r^{\mathcal{I}^r_\mathcal{O}}\} \neq \emptyset$ has an implicant $r \in \Gamma$. Lemma 4.5 then yields:

- $\tau(t) = \tau(t_\zeta)$ for all $t \in \zeta$, and
- $\tau(t) = \tau(t^{\mathsf{Pre}}_\zeta)$ for all $t \in \mathsf{Pre}(\zeta)$.

Let $\Psi := \{s \in \mathsf{N_R} \mid s(t, t') \in \phi$ for some $t \in \mathsf{Pre}(\zeta), t' \in \zeta\}$. Then $\Psi \subseteq \Gamma$ and hence r is an implicant for Ψ; moreover, there exists a prime implicant $\hat{r} \in \Gamma$ of Ψ. Then we have $(\tau(t^{\mathsf{Pre}}_\zeta), \tau(t_\zeta)) \in \hat{r}^{\mathcal{I}^r_\mathcal{O}}$ and $\hat{r} \in \mathsf{I}$. $\square$

The following is a direct consequence of Lemmas 4.4 and 4.6.

Corollary 4.7. *Let $\phi = \exists \vec{x}.\psi$ be a CQ. If $\mathcal{U}_\mathcal{O} \models \psi(a_1, \ldots, a_k)$, then $\mathcal{I}^r_\mathcal{O} \models \phi^\dagger_\mathcal{R}(a_1, \ldots, a_k)$.*

To finish the proof of Theorem 4.1, we need only to show that the converse implication holds too.

Lemma 4.8. *If $\mathcal{I}^r_\mathcal{O} \models \phi^\dagger_\mathcal{R}(a_1, \ldots, a_k)$, then $\mathcal{U}_\mathcal{O} \models \phi(a_1, \ldots, a_k)$.*

Proof. Let π be an $(a_1, \ldots, a_k)$-match for $\mathcal{I}^r_\mathcal{O}$ and $\phi^\dagger_\mathcal{R}$. We start with introducing some notation. The degree $d(\zeta)$ of an equivalence class ζ is the length $n \geq 0$ of a longest sequence (if it exists) $r_0(t_0, t'_0), \ldots, r_n(t_n, t'_n) \in \phi$ such that $r_i \in \mathsf{N_R}$, $t_0 \in \zeta$ and $t'_i \sim_{\phi^r} t_{i+1}$ for all $i < n$. If no longest sequence exists, we set $d(\zeta) = \infty$.

Claim 1.
(a) If $\pi(t) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$*, then* $d([t]_{\sim_{\phi^r}}) < \infty$.

(b) If $s \sim_{\phi^r} t$ *and* $\pi(s) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$*,then*

(i) $\pi(s) = \pi(t)$;

(ii) If $r_1(s', s), r_2(t', t) \in \phi$, $r_1, r_2 \in \mathsf{N_R}$*, then* $\pi(s') = \pi(t')$.

We start with proving (a). Assume to contrary of what has to be shown that there is a t_0 with $\pi(t_0) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ and an infinite sequence $r_0(t_0, t'_0), r_1(t_1, t'_1), \ldots$ with $r_i \in \mathsf{N_R}$ and $t'_i \sim_{\phi^r} t_{i+1}$ for all $i \geq 0$. By definition of $(a_1, \ldots, a_k)$-match, $\pi(t_0) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ implies that $t_0 \in \mathsf{QVar}(\phi)$. As ϕ is finite, there exist m, n with $0 \leq m \leq n$ such that $t'_n = t'_m$. It follows that $t_0 \in \mathsf{RCyc}$. Hence ψ_1 contains the conjunct $\neg\mathsf{Aux}(t_0)$ and we have derived a contradiction to $\pi(t_0) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$.

We now consider (b). Because of (a), Point (i) of (b) can be proved by induction on $n := d([s]_{\sim_{\phi^r}}) = d([t]_{\sim_{\phi^r}})$. For the induction start, let $s \sim_{\phi^r} t$ with $\pi(s) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ and $d([s']_{\sim_{\phi^r}}) = 0$. By definition of $\sim_{\phi^r}$, we have $s \sim_{\phi} t$. If $s = t$, $\pi(s) = \pi(t)$ trivially holds. In case $s \neq t$, the definition of $\sim_{\phi^r}$ yields that we have $r_1(s', s), r_2(t', t) \in \phi, r_1, r_2 \in \mathsf{N_R}$, and hence $\pi(t) \notin \mathsf{N}_{\mathsf{I}}^{\rho}$, , by the definition of $\mathcal{I}^r_{\mathcal{O}}$. Therefore, $\pi(s) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ and $s \sim_{\phi} t$ lead to $\pi(s) = \pi(t)$, by Lemma 3.4. For the induction step, define

$$
\begin{aligned}
\sim^{(0)}_{\phi^r} &:= \{(t, t) \mid t \in \mathsf{Term}(\phi)\} \cup \\
&\quad\ \{(t, t') \mid r_1(s, t), r_2(s', t') \in \phi, r_1, r_2 \in \mathsf{N_R}, t \sim_{\phi} t'\} \\
\sim^{(i+1)}_{\phi^r} &:= \sim^{(i)}_{\phi^r} \cup \\
&\quad\ \{(s, t) \mid \text{there is } s' \text{ with } s \sim^{(i)}_{\phi^r} s' \text{ and } s' \sim^{(i)}_{\phi^r} t\} \cup \\
&\quad\ \{(s, t) \mid \text{there are } r_1(s, s'), r_2(t, t') \in \phi, r_1, r_2 \in \mathsf{N_R}, s' \sim^{(i)}_{\phi^r} t'\}
\end{aligned}
$$

for all $i \geq 0$. It is not hard to see that $\sim_{\phi^r} = \bigcup_{i \geq 0} \sim^{(i)}_{\phi^r}$. We show by induction on i that if $s \sim^{(i)}_{\phi^r} t$, $d([s]_{\sim_{\phi^r}}) = n$, and $\pi(s) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$, then $\pi(s) = \pi(t)$. The induction start is trivial for the identity part of $\sim^{(0)}_{\phi^r}$, which implies $s = t$. For the other part, the arguments are the same as those given for the start of the outer induction. For the induction step, we distinguish two cases:

- There is s' with $s \sim^{(i)}_{\phi^r} s'$ and $s' \sim^{(i)}_{\phi^r} t$. By (inner) IH, $\pi(s) = \pi(s')$ and thus $\pi(s') \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$. Since $s \sim^{(i)}_{\phi^r} s'$, we have $[s]_{\sim_{\phi^r}} = [s']_{\sim_{\phi^r}}$, , and thus $d([s]_{\sim_{\phi^r}}) = n$. We hence can apply (inner IH) once more to derive $\pi(s') = \pi(t)$ and thus get $\pi(s) = \pi(t)$.

- There are $r_1(s, s'), r_2(t, t') \in \phi$ such that $s' \sim^{(i-1)}_{\phi^r} t'$. By the definition of $\mathcal{I}^r_{\mathcal{O}}$, $r(s, s') \in \phi$ and $\pi(s) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ entails $\pi(s') \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$. By definition of depth, we further have $d([s']_{\sim_{\phi^r}}) < d([s]_{\sim_{\phi^r}})$. We thus can apply (outer) IH to obtain $\pi(s') = \pi(t_{[s']_{\sim_{\phi^r}}})$. Hence, $\pi(t_{[s']_{\sim_{\phi^r}}}) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$. Thus, from the conjunct ψ_2 of $\phi^{\dagger}_{\mathcal{R}}$, we obtain $\pi(s) = \pi(t)$.

Now for Point (ii). Assume $\pi(s) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$, $r1(s', s), r2(t', t) \in \phi$, and $s \sim_{\phi^r} t$. By Point (i), $\pi(s) = \pi(t_{[s]_{\sim_{\phi^r}}})$. Hence, by the conjunct ψ_2 of $\phi^{\dagger}_{\mathcal{R}}$, and the fact that π is a match for $\phi^{\dagger}_{\mathcal{R}}$ and $\mathcal{I}^r_{\mathcal{O}}$, $\pi(s') = \pi(t')$. This finishes the proof of Claim 1.

Let $\sim_\pi$ be the transitive closure of

$$
\begin{aligned}
&\{(t,t) \mid t \in \mathsf{Term}(\phi)\} \cup \\
&\{(s,t) \in \mathsf{Term}(\phi) \times \mathsf{Term}(\phi) \mid s \sim_{\phi^r} t, \pi(s), \pi(t) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}\} \cup \\
&\{(s,t) \in \mathsf{Term}(\phi) \times \mathsf{Term}(\phi) \mid \exists r_1, r_2 \in \mathsf{N_R}, r_1(s,s'), r_2(t,t') \in \phi : \pi(s') \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}, \\
&s' \sim_{\phi^r} t'\}.
\end{aligned}
$$

By Claim 1, we have
(*) $\pi(s) = \pi(t)$ whenever $s \sim_\pi t$. Note that $\sim_\pi$ is an equivalence relation because it is, by Claim 1, the transitive closure of a symmetric relation. Now let the query ϕ' be obtained from $\exists\vec{x}.\vec{y}.\psi_{\mathsf{CQ}}$ by identifying all terms $t, t' \in \mathsf{Term}(\phi)$ such that $t \sim_\pi t'$. More precisely, choose from each $\sim_\pi$-equivalence class ξ a fixed term $t_\xi \in \xi$ and replace each occurrence of an element of ξ in $\exists\vec{x}.\vec{y}.\psi_{\mathsf{CQ}}$ by t_ξ. By (*), π is a match for $\mathcal{I}^r_{\mathcal{O}}$ and the resulting query ϕ'.

Claim 2. *The unfolding of ϕ, has the following properties:*

(a) For all variables $v \in \mathsf{Var}(\exists\vec{x}.\vec{y}.\psi_{\mathsf{CQ}}) \setminus \mathsf{Var}(\phi)$ there is maximally one atom $r(t, v) \in \phi^{\dagger}_{\mathcal{R}}$, $r \in \mathsf{N_R} \cup \{\rho_L\}$, $t \in \mathsf{Term}(\phi^{\dagger}_{\mathcal{R}})$.

(b) For all variables $v \in \mathsf{Var}(\phi)$, we have that if $r(t, v) \in \phi^{\dagger}_{\mathcal{R}}$, $r \in \mathsf{N_R} \cup \{\rho_L\}$, $t \in \mathsf{Term}(\phi^{\dagger}_{\mathcal{R}})$, then $r(t, v) \in \phi$.

(c) If there is a sequence $r_0(t_0, t'_0), \ldots, r_m(t_m, t'_m) \in \phi^{\dagger}_{\mathcal{R}}$ with $m \geq 0$, $t'_i \sim_{\phi^r} t_{i+1}$ or $t'_i \sim_\phi t_{i+1}$, for all $i < m$, and $t'_m \sim_\phi t_0$, then $t_i, t'_i \notin \mathsf{Var}(\exists\vec{x}.\vec{y}.\psi_{\mathsf{CQ}}) \setminus \mathsf{Var}(\phi)$ and especially $r_i \neq \rho_L$, for all r_i.

Note that the unfolding can be considered inductively. Moreover, it can be easily seen that each unfolding step uses a freshly introduced variable maximally once as successor in an atom $r \in \mathsf{N_R} \cup \{\rho_L\}$ that is introduced in the same step and that it does not use other than such variables as successors in such atoms. Note that this directly implies (a) and (b). Together with the fact that the unfolding only uses fresh variables as successors, the assumption that a predicate $\rho_L \in \phi^{\dagger}_{\mathcal{R}}$ can only have been introduced during unfolding yields (c).

We now can show the following:

(I) If $v \in \mathsf{QVar}(\phi')$ with $\pi(v) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$, then there is at most one $t \in \mathsf{Term}(\phi')$ such that $r(t, v) \in \phi'$, for some $r \in \mathsf{N_R} \cup \{\rho_L\}$;

(II) If $v \in \mathsf{QVar}(\phi')$ with $\pi(v) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ and $t \in \mathsf{Term}(\phi')$ such that $\Gamma = \{r \mid r(t, v) \in \phi'\} \neq \emptyset$, then there is an implicant s for Γ with $(\pi(t), \pi(v)) \in s^{\mathcal{I}^r_{\mathcal{O}}}$;

(III) If $r_0(t_0, t'_0), \dots, r_m(t_m, t'_m) \in \phi'$ with $m \geq 0$, $r_i \in \mathsf{N_R} \cup \{\rho_L\}$, $t'_i \sim_\phi t_{i+1}$, for all $i < m$, and $t'_m \sim_\phi t_0$, then $\pi(t_i), \pi(t'_i) \notin \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$, for all $i \leq m$.

First for (I). Let $\pi(v) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ and $r_1(t_1, v), r_2(t_2, v) \in \phi'$, $r_1 \neq r_2$. By Claim 2(a), we then have that v cannot be a variable introduced during unfolding. Thus, Claim 2(b) and the construction of $\exists\vec{x}.\vec{y}.\psi_{\mathsf{CQ}}$ yields that $r_1(s_1, s'_1), r_2(s_2, s'_2) \in \phi$ such that $s_1 \sim_\pi t$, $s_2 \sim_\pi t$, and $s_1 \sim_\pi v \sim_\pi s_2$. By (*), $\pi(s_1) = \pi(v)$, and thus $\pi(s') \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$. By definition of $\sim_\pi$, $s' \sim_\pi s'$ implies $s'_1 \sim_{\phi^r} s'_2$. Summing up, we thus have $t_1 \sim_\pi t_2$. Since both t_1 and t_2 occur in ϕ', we have $t_1 = t_2$.

Now for (II). By Claim 2(a), we obviously have such an implicant for all variables introduced during unfolding, which is, if it exists, the unique role atom in which such a variable occurs as successor. Let now $v \in \mathsf{Var}(\phi)$, $\pi(v) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$, and $\Gamma \neq \emptyset$. Due to the use of $\mathsf{Fork}_{\neq}$ in ψ_1 and since $\pi(v) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$, there is an implicant for $\mathsf{In}([v]_{\sim_{\phi^r}})$. By ψ_3, there thus is an implicant s for $\mathsf{In}([v]_{\sim_{\phi^r}})$ with $(\pi(t^{\mathsf{Pre}}_{[v]}), \pi(t_{[v]})) \in s^{\mathcal{I}^r_{\mathcal{O}}}$. Since $\pi(v) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ we have $t^{\mathsf{Pre}}_{[v]} \sim_\pi t_{[v]}$ and $t_{[v]} \sim_\pi v$. By (*), $\pi(t_{[v]}) = \pi(t)$ and $\pi(t_{[v]}) = \pi(v)$, thus $(\pi(t), \pi(v)) \in s^{\mathcal{I}^r_{\mathcal{O}}}$. Since $\Gamma \subseteq \mathsf{In}([v]_{\sim_{\phi^r}})$, s is the required implicant for Γ.

For (III), let $r_0(t_0, t'_0), \dots, r_m(t_m, t'_m) \in \phi'$ with $m \geq 0$, $r_i \in \mathsf{N_R} \cup \{\rho_L\}$, $t'_i \sim_\phi t_{i+1}$, for all $i < m$, and $t'_m \sim_\phi t_0$. By the construction of $\exists\vec{x}.\vec{y}.\psi_{\mathsf{CQ}}$, which does not replace variables introduced during unfolding, and Claim 2(c), we have that there are $r_0(s_0, s'_0), \dots, r_m(s_m, s'_m) \in \phi$ with $s_i \sim_\pi t_i$ and $s'_i \sim_\pi t'_i$ and for all $i \leq m$, $s'_i \sim_\phi s_{i+1}$, for all $i < m$, and $s'_m \sim_\phi s_0$. Assume now, contrary to what has to be shown, that $\pi(t'_i) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ for some $i \leq m$. Since $s_i \sim_\pi t_i$, (*) yields $\pi(s) = \pi(t)$. Thus $\pi(s_i) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$, which implies $s_i \in \mathsf{QVar}(\phi)$ by definition of $(a_1, \dots, a_k)$-matches. Together with $\sim_\pi \subseteq \sim_{\phi^r}$, $s_i \in \mathsf{QVar}(\phi)$ implies $s_i \in \mathsf{Cyc}$. Thus, $\neg\mathsf{Aux}(s_i)$ is a conjunct of ψ_1 and $\pi(s) \notin \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$, which is a contradiction. This finishes the proof of (I)-(III).

We now inductively define a mapping $\tau : \mathsf{Term}(\phi') \to \Delta^{\mathcal{U}_{\mathcal{O}}}$ such that $\mathsf{Tail}(\tau(t)) = \pi(t)$ for all $t \in \mathsf{Term}(\phi')$ and $(\tau(t), \tau(v)) \in \rho^{\mathcal{U}_{\mathcal{O}}}$ if $(\pi(t), \pi(v)) \in \rho^{\mathcal{I}^r_{\mathcal{O}}}$. . For the induction start, we consider three cases:

(i_0) For all $t \in \mathsf{Term}(\phi')$ with $\pi(t) \notin \mathsf{Aux} \cup \mathsf{Aux}_\rho$, set $\tau(t) := \pi(t)$. Note that this defines $\tau(t)$ for all $t \in \mathsf{AVar}(\phi') \cup (\mathsf{Term}(\phi') \cap \mathsf{N_I})$.

(ii_0) For all $v \in \mathsf{QVar}(\phi')$ with $\pi(v) \in \mathsf{Aux}^{\mathcal{I}^r_{\mathcal{O}}}$ and such that there is neither an atom $r(t, v) \in \phi_L$, $r \in \mathsf{N_R} \cup \{\rho_L\}$, nor a symbol $t \in \mathsf{Term}(\phi')$ with $v \sim_{\phi^r} t$ and $v \neq t$ (i.e., there is no atom $\rho(v, t') \in \phi$ or $\rho(t', v) \in \phi$, $t' \in \mathsf{Term}(\phi')$), do the following. By the definition of $\mathcal{U}_\mathcal{O}$ and because each $d \in \Delta^{\mathcal{I}^r_\mathcal{O}}$ is reachable from an element of $\mathsf{Ind}(\mathcal{A})^{\mathcal{I}^r_\mathcal{O}}$, there is a sequence $d_0, \ldots, d_n \in \Delta^{\mathcal{I}^r_\mathcal{O}}$ and a sequence $r_0, \ldots, r_{n-1} \in \mathsf{N_R} \cup \{\rho\}$ such that $d_0 \in \mathsf{Ind}(\mathcal{A})^{\mathcal{I}^r_\mathcal{O}}$,$d_n = \pi(v)$, $(d_i, d_{i+1}) \in r^{\mathcal{I}^r_\mathcal{O}}$ if $r \in \mathsf{N_R}$, and $(d_i, d_{i+1}) \in \rho_\mathcal{O}$ if $r = \rho$ for all $0 \leq i < n$. Set $\tau(v) := d_0 r_0 d_1 \cdots r_{n-1} d_n \in \Delta^{\mathcal{U}_\mathcal{O}}$.

(iii$_0$) For all $v \in \mathsf{QVar}(\phi')$ with $|[v]_{\sim_{\phi^r}}| > 1$, for which we have that all $t \in \mathsf{Term}(\phi')$ with $(v, t) \in \sim_{\phi^r}$ are such that $\tau(t)$ is not defined yet and there is no atom $r(t', t) \in \phi$, $r \in \mathsf{N_R} \cup \{\rho_L\}$, $t' \in \mathsf{Term}(\phi')$, proceed as in (ii$_0$).

For the induction step, proceed as follows.

(i) If $\tau(v)$ is undefined and there exists an atom $r(t, v) \in \phi'$ with $r \in \mathsf{N_R}$ and $\tau(t)$ defined, then (II) yields an implicant s for $\Gamma = \{r \mid r(t, v) \in \phi'\} \neq \emptyset$ with $(\pi(t), \pi(v)) \in s^{\mathcal{I}^r_\mathcal{O}}$. Set $\tau(v) := \tau(t) \cdot s\pi(v)$. Since $\mathsf{Tail}(\tau(t)) = \pi(t)$ and $(\pi(t), \pi(v)) \in s^{\mathcal{I}^r_\mathcal{O}}$, we have $\tau(v) \in \Delta^{\mathcal{U}_\mathcal{O}}$.

(ii) If $\tau(v)$ is undefined and there exists a symbol $t \in \mathsf{Term}(\phi')$ with $v \sim_{\phi^r} t$ and $\tau(t)$ defined, do the following.

   (a) If $\pi(t) \in [a]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$, $a \in \mathsf{Ind}(\mathcal{A})$, set $\tau(v)$ to an arbitrary element $p \in \Delta^{\mathcal{U}_\mathcal{O}}$ with $\mathsf{Tail}(p) = \pi(v)$.

   (b) If $\pi(t) \in [x_C]_{\rho^{\mathcal{I}^r_\mathcal{O}}}$, $x_C \in \mathsf{N_I^{aux}}$, we apply that $\mathsf{Tail}(\tau(t)) = \pi(t)$, given by construction. By Lemma 3.4 and Proposition 3.7, we then get that $\tau(t)$ is of the form $\tau(t) = p \cdot rx_C$, $\tau(t) = p \cdot rx_C(\rho x_{D'_i,x_C})^i$, or $\tau(t) = p \cdot rx_C\rho a_{x_C}$ for some $r \in \mathsf{N_R}$, $p \in \Delta^{\mathcal{U}_\mathcal{O}}$, and $i \geq 1$. If $\pi(v) = x_C$ set $\tau(v) := p \cdot rx_C$. If $\pi(v)$ is of the form $\pi(v) = x_{E,x_C}$, then the definition of $\mathcal{U}_\mathcal{O}$, and Proposition 3.7 yield that we have an element $p' \cdot x_C(\rho x_{E'_j,x_C})^j \rho x_{E,x_C} \in \Delta^{\mathcal{U}_\mathcal{O}}$, $j \geq 0$. But then, we also have the element $e = p \cdot rx_C(\rho x_{E'_j,x_C})^j \rho x_{E,x_C} \in \Delta^{\mathcal{U}_\mathcal{O}}$, and can set $\tau(v) := e$. The case for $\pi(v) = a_{x_C}$ is analogous to the previous case.

(iii) If $\tau(v)$ is undefined and there exists an atom $\rho_L(t, v) \in \phi'$ with $\tau(t)$ defined, then set $\tau(v) := \tau(t) \cdot \rho\pi(v)$. Since $\mathsf{Tail}(\tau(t)) = \pi(t)$ and $(\pi(t), \pi(v)) \in \rho_L^{\mathcal{I}^r_\mathcal{O}}$, we have $\tau(v) \in \Delta^{\mathcal{U}_\mathcal{O}}$.

The mapping τ is clearly well-defined for the three base cases of the induction. For (iii$_0$) this is the case because it always is applicable only once for an equivalence class of $\sim_\phi$. By (I), τ is well-defined for the first induction step (i.e., the term t in the induction step is unique). For (ii) we must show that if there are several t already defined, the equivalence class chosen for $\tau(v)$ is the same for all t. This is obviously the case if we have $\pi(t) \in \mathsf{Ind}(\mathcal{A})$ for any such t. If this is not the case, $\tau(t)$ must have been defined by (a) (iii$_0$) or (b) during induction. For (a), we have that (iii$_0$) can only be applied for one t and only if there is no atom

$r(t', t'') \in \phi'$, with $t', t'' \in \mathsf{Term}(\phi')$, $r \in \mathsf{N_R} \cup \{\rho_L\}$ and $t \sim_\phi t''$. In the subsequent induction to define $\tau(t)$ for all other t, hence only step (ii) is applicable and the equivalence class chosen for $\tau(v)$ is the one selected during application of (iii$_0$). For (b), Claim 2(b) yields that if some $\tau(t)$ is defined during induction by step (i) or (iii), this can only happen once and only if (iii$_0$) was not applied before. Hence, in case the equivalence class of $\tau(t)$ was chosen 'arbitrarily' (i.e., though in 'Tail-accordance' to $\pi(t)$) there is no other such choice in future for any $t' \in \mathsf{Term}(\phi')$ with $t \sim_\phi t''$. Step (iii) is well-defined because atoms of the form $\rho_L(t, v) \in \phi'$ must have been introduced during the construction of ψ_{CQ}, which does not use a variable as v twice as successor (i.e., again, the term t in the induction step is unique). By construction of ϕ_L, we further have that if the third induction step can be applied to an atom $\rho_L(t, v)$, then (i) cannot be applicable, by I. The latter also yields that only one (i) or (iii) can be applicable with (ii) for a v at the same time. But then, there must be some t defined, which can only have happened by (i$_0$). Hence, there is no ambiguity with the selection of the equivalence class.

Moreover, the mapping τ is total, which means that $\tau(t)$ is defined for all $t \in \mathsf{Term}(\phi')$. This is partly because of (III), which describes that there cannot be a 'cycle' of role/ρ atoms in ϕ' where one variable is mapped to an unnamed element (i.e., such a variable is potentially undefined). Hence, the mapping is defined for the terms occurring in such a cycle. In addition, the case of a cycle formed by only ρ atoms is covered in base case (iii) of the definition of τ.

The constructed τ is also a match for $\mathcal{U}_\mathcal{O}$ and ϕ'. To show that, we only need to consider concepts of the form $A \in \mathsf{N_C}$ because of the unfolding. It is immediate that $\mathcal{U}_\mathcal{O} \models \tau A(t)$ for all $A(t) \in \phi'$ by $\mathsf{Tail}(\tau(t)) = \pi(t)$, which is a property of the construction of τ, and Lemma 3.11. Now let $r(t, t') \in \phi'$, $r \in \mathsf{N_R}$. If $\pi(t), \pi(t') \notin \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}} \cup \mathsf{Aux}_\rho^{\mathcal{I}^r_\mathcal{O}}$, then $\tau(t) = \pi(t), \tau(t') = \pi(t')$, and $(\pi(t), \pi(t')) \in r^{\mathcal{U}_\mathcal{O}}$, by the definition of $\mathcal{U}_\mathcal{O}$. If $\pi(t') \in \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}}$, then the construction of τ implies that $\tau(t') = \tau(t) \cdot s\pi(t)$ with $\mathcal{T} \models s \sqsubseteq r$. By the definition of $\mathcal{U}_\mathcal{O}$, it follows that $(\tau(t), \tau(t')) \in r^{\mathcal{U}_\mathcal{O}}$. The cases that $\pi(t) \in \mathsf{Aux}^{\mathcal{I}^r_\mathcal{O}} \cup \mathsf{Aux}_\rho^{\mathcal{I}^r_\mathcal{O}}$ and $\pi(t') \in \mathsf{Ind}(\mathcal{A})$, and $\pi(t') \in \mathsf{Aux}_\rho^{\mathcal{I}^r_\mathcal{O}}$ cannot occur, by the definition of $\mathcal{I}^r_\mathcal{O}$. For $\rho(t, t') \in \phi'$, $(\pi(t), \pi(t')) \in \rho^{\mathcal{I}^r_\mathcal{O}}$, given by the semantics, directly yields that $(\tau(t), \tau(t')) \in \rho^{\mathcal{U}_\mathcal{O}}$ since this is a property of the construction of τ. For $\rho_L(t, t') \in \phi'$, we have $(\pi(t), \pi(t')) \in \rho_L^{\mathcal{I}^r_\mathcal{O}}$ and that $\pi(t)$ and $\pi(t')$ must be of the form e and a_e, $e \in \mathsf{Ind}(\mathcal{A}) \cup (\mathsf{N_I^{aux}} \cap \Delta^{\mathcal{I}^r_\mathcal{O}})$. But then, the construction of τ implies that there is an element $\tau(t') = \tau(t) \cdot \rho\pi(t') \in \Delta^{\mathcal{U}_\mathcal{O}}$, and then the definition of $\rho_L^{\mathcal{U}_\mathcal{O}}$ yields $(\tau(t), \tau(t')) \in \rho_L^{\mathcal{U}_\mathcal{O}}$.

Finally, we adapt τ to get a mapping from $\mathsf{Term}(\phi)$ to $\Delta^{\mathcal{U}_\mathcal{O}}$ by setting $\tau(t) := \tau(t')$ if $t \in \mathsf{Term}(\phi) \setminus \mathsf{Term}(\phi')$ and $t \sim_{\phi^r} \pi(t')$. It is straightforward to verify that τ is a match for $\mathcal{U}_\mathcal{O}$ and ϕ. Since $\tau(t) = \pi(t)$ if $\pi(t) \in \mathsf{Ind}(\mathcal{A})$ for all $t \in \mathsf{Term}(\phi')$, it is also clear that τ is an $(a_1, \ldots, a_k)$-match.

□

5 Conclusions

In this report, we proposed an extension of the combined approach of answering CQs over $\mathcal{ELH}_\perp$-ontologies [LTW09] to answer CQs over ontologies in the rough DL $\mathcal{ELH}_{\perp\rho}$. This, in particular, consists of the extension of the classical canonical model construction as well as that of the definition of the query rewriting. Since our techniques especially retain the P complexity of the original approach, we provide a method to directly address vague knowledge in ontologies as well as CQs that does not further increase the complexity.

References

[BCM+07] 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, 2nd edition, 2007.

[LTW09] Carsten Lutz, David Toman, and Frank Wolter. Conjunctive query answering in the description logic el using a relational database system. In *IJCAI*, pages 2070–2075, 2009.

[PZ13] Rafael Peñaloza and Tingting Zou. Roughening the $\mathcal{EL}$ envelope. In P. Fontaine, C. Ringeissen, and R. A. Schmidt, editors, *Proceedings of the 2013 International Symposium on Frontiers of Combining Systems (FroCoS 2013)*, volume 8152 of *Lecture Notes in Computer Science*, pages 71–86, Nancy, France, 2013. Springer-Verlag.

[SKP07] Stefan Schlobach, Michel C. A. Klein, and Linda Peelen. Description logics with approximate definitions - precise modeling of vague concepts. In Manuela M. Veloso, editor, *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI-07)*, pages 557–562, 2007.