Query Answering for Rough $\mathcal{EL}$ Ontologies

Rafael Peñaloza
KRDB Research Centre,
Free University of Bolzano, Italy
Rafael.Penaloza@unibz.it

Veronika Thost
MIT-IBM Watson AI Lab
IBM Research
veronika.thost@ibm.com

Anni-Yasmin Turhan
Inst. of Theor. Computer Science
TU Dresden, Germany
Anni-Yasmin.Turhan@tu-dresden.de

Abstract

Querying large datasets with incomplete and vague data is still a challenge. Ontology-based query answering extends standard database query answering by background knowledge from an ontology to augment incomplete data. We focus on ontologies written in rough description logics (DLs), which allow to represent vague knowledge by partitioning the domain of discourse into classes of indiscernible elements.

In this paper, we extend the combined approach for ontology-based query answering to a variant of the DL $\mathcal{ELH}_\bot$ augmented with rough concept constructors. We show that this extension preserves the good computational properties of classical $\mathcal{EL}$ and can be implemented by standard database systems.

Introduction

Ontology-based query answering performs database-style query answering over description logic (DL) knowledge bases (KBs), which consist of an ontology (or TBox) expressing terminological (i.e., background) knowledge about a domain, and a dataset (called ABox) containing facts about particular individuals. The knowledge in the KB is captured by means of concepts (unary predicates) and roles (binary relations). The use of conceptual background knowledge allows one to derive more answers to queries than from the data alone. The queries considered are typically conjunctive queries, which are special forms of first-order (FO) queries. The expressivity of a DL is determined by the concept (and sometimes also role) constructors it provides to describe important notions from the application domain. In classical DLs concepts represent unary predicates and hence are interpreted as sets of elements. Thus, classical DLs lack capabilities of modeling uncertainty or vagueness (Lukasiewicz and Straccia 2008).

A moderate form of relaxation of concepts can be achieved by interpreting them as *rough sets* (Pawlak 1982). Rough sets employ an *indiscernibility relation* ρ, which groups objects that are considered to be indistinguishable from one another. The relation ρ effectively partitions the set of elements into so-called *granules*. A granule, in essence, relaxes the notion of an element to a class of equivalent elements. In rough sets, every classic set, say S, is accompanied by two sets. The *lower approximation* $\underline{S}$ contains elements that all share the properties of elements in S as it contains those partitions that lie completely in S. The *upper approximation* $\overline{S}$ contains elements that are indistinguishable from an element in S, i.e., it contains those granules that overlap with S. Rough sets are employed in knowledge discovery and data mining, among others (Lin and Cercone 2012). The capability of rough sets to relax objects in the data was already noticed in (Pawlak 1998) and is a standard way to relax database queries. One of the goals of this paper is to extend these ideas to relax ontology-based query answering techniques.

In the context of DLs, concept constructors for upper (and lower) approximations provide means to relax (and crispen) concepts, while granules effectively relax objects. The idea to use rough set interpretations for DLs is not new (Liau 1996; Klein, Mika, and Schlobach 2007; Schlobach, Klein, and Peelen 2007; Jiang et al. 2009; Keet 2010). Rough DLs typically have concept constructors for the upper and the lower approximation of concepts. One of their basic motivations is medical applications (Klein, Mika, and Schlobach 2007; Schlobach, Klein, and Peelen 2007), where, for instance, patients can be indistinguishable by their symptoms or drugs and their generica can be indistinguishable by their active agent. Similarly, they were suggested to enhance the web ontology language OWL (Keet 2010) or to solve the identity matching problem in the linked data cloud (Klein, Mika, and Schlobach 2007; Beek, Schlobach, and van Harmelen 2016). As in database settings, indiscernibility relations for rough DLs can be derived automatically from the data (d'Amato et al. 2013; Beek, Schlobach, and van Harmelen 2016) making rough DLs amenable for practical applications.

Another approach for dealing with vagueness is based on fuzzy logic. While fuzzy DLs (Bobillo et al. 2015) can express vagueness regarding the concept membership of objects, rough DLs can express granularity of objects. The former DLs can easily turn undecidable (Borgwardt, Distel, and Peñaloza 2015; Borgwardt, Cerami, and Peñaloza 2017), but the latter are always decidable, as long as the underlying classical DL is. Reasoning procedures for classical reasoning tasks such as satisfiability or subsumption, i.e., the computation of sub- and super-concept relationships in rough DLs were proposed in (Klein, Mika, and Schlobach 2007;

Keet 2011; Peñaloza and Zou 2013). In fact, if inverse roles, transitive roles and role hierarchies are available in a DL, then reasoning in its rough variant can be reduced to it (Klein, Mika, and Schlobach 2007). The lightweight DL $\mathcal{EL}$ has only conjunction and existential restrictions as concept constructors and thus such a reduction would use a much more expressive logic with higher computational complexity. $\mathcal{EL}$ cannot express contradictions, thus subsumption is the interesting reasoning task, and can be decided in polynomial time (Baader, Brandt, and Lutz 2005) by means of canonical models (Lutz and Wolter 2010). The subsumption decision procedure based on canonical models was lifted in (Peñaloza and Zou 2013) to $\mathcal{ELH}^{\rho}_{\bot}$—a rough variant of $\mathcal{EL}$ with role hierarchies extended by constructors for upper and lower approximations of concepts. This rough DL can be used, for example, to model biological species through their phenotypical characteristics, which are often vague in nature. For example, the edible *Agaricus arvensis* mushroom is described to have an 'anise-like' smell, 'ellipsoid' spores, among other characteristics. Thus, we can say that this mushroom belongs to the concept

$$\mathsf{Edible} \sqcap \exists \mathsf{hasSmell}.\overline{\mathsf{Anise}} \sqcap \exists \mathsf{hasSpores}.\exists \mathsf{hasShape}.\underline{\mathsf{Ellipse}}.$$

We consider ontology-based query answering in the DL $\mathcal{ELH}^{\rho}_{\bot}$. For this task, we use conjunctive queries that admit $\mathcal{ELH}^{\rho}_{\bot}$ concepts and the indiscernibility relation ρ in the atoms of the query. For example, when preparing a field-guide to mushroom picking, it is important to highlight possible confusions between edible and poisonous mushrooms to avoid an intoxication. More precisely, one could query for all pairs of mushrooms that are morphologically similar, but where one is edible and the other is not, through the query

$$\begin{aligned}\Phi(x_1,x_2) = \exists y_1,y_2.&\mathsf{Mushroom}(x_1) \wedge \mathsf{Edible}(x_1) \wedge\\ &\mathsf{Mushroom}(x_2) \wedge \mathsf{Poisonous}(x_2) \wedge\\ &\mathsf{hasShape}(x_1,y_1) \wedge\\ &\mathsf{hasShape}(x_2,y_2) \wedge \rho(y_1,y_2).\end{aligned}$$

Such a query can be further refined, for example, to return additionally the smell of the poisonous elements, or to consider other characteristics like color, size, or the shape of the spores. In this case, the query described above could return the two answers that *Agaricus arvensis* (which is edible) may be confused with the poisonous *Agaricus xanthodermus* and with *Agaricus pilatianus*. The refined query would state that both poisonous species have a pungent smell, which makes them easy to differentiate from *Agraricus arvensis*.

Obviously, the relevance of rough CQ answering is not limited to the identification of mushrooms or other biological species. It has also applications in medicine (Schlobach, Klein, and Peelen 2007), for suggesting adequate treatments after identifying symptoms, and diseases, which usually have vague descriptions. Furthermore rough CQ answering is applied in verification, for quality control; and in online marketing, for handling similar clients uniformly, among many others.

A well-known approach to answering conjunctive queries for classical $\mathcal{EL}$ is the *combined approach* (Lutz, Toman, and Wolter 2009). It proceeds in two steps. First, all the knowledge from the TBox is 'absorbed' into the ABox. After this step only the data in the materialized ABox, but not the TBox, needs to be regarded for answering the query. The materialized ABox introduces auxiliary elements to represent information about all syntactical sub-concepts occurring in the TBox. Hence, such a materialized ABox may give 'spurious' answers to the original query, due to joins at auxiliary elements in the materialized ABox. In the second step of the approach, the query is rewritten. The rewriting complements the query with filter conditions that sift out the spurious answers. The combined approach is designed to be implemented by database systems. The materialized ABox can be represented in a database and the rewritten conjunctive query can be expressed by standard database query languages. This approach has been implemented in competitive systems such as Combo system (Lutz et al. 2013), and, based on Datalog, in RDFox (Motik et al. 2014) and HermiT (Stefanoni and Motik 2015).

To lift the combined approach for $\mathcal{EL}$ to the rough DL $\mathcal{ELH}^{\rho}_{\bot}$ the materialization and the rewriting, both need to be extended. The materialized ABox needs to be further augmented by new auxiliary elements. These new elements represent the upper and lower approximations of concepts. Due to their semantics, they can give rise to new kinds of joins, which can in turn cause new kinds of spurious elements. These new joins are not detected by the filters employed for the classical $\mathcal{EL}$ query answering method. Thus, it is important to provide new filter predicates for the rewritten query in the presence of rough information.

For lack of space, full proofs are not included in this paper. They can be found in the technical report (Peñaloza, Thost, and Turhan 2018).

Preliminaries

We introduce the rough DL $\mathcal{ELH}^{\rho}_{\bot}$, that extends the classical DL $\mathcal{ELH}_{\bot}$ by an indiscernibility relation and by concept constructors for the lower and the upper approximation. Based on this, we define the problem of answering conjunctive queries that we consider.

Syntax. 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*, respectively, and let ρ be the *indiscernibility relation*. $\mathcal{ELH}^{\rho}_{\bot}$ *concepts* are built inductively by the following syntax rule (where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$):

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

Concepts of the form $\overline{C}$ (resp. $\underline{C}$) are called the *upper* (resp. *lower*) *approximation* of C. Let $A \in \mathsf{N_C}$, $r,s \in \mathsf{N_R}$, $a,b \in \mathsf{N_I}$, and C and D be concepts. *Axioms* are the following kinds of expressions: *general concept inclusions* (GCIs) of the form $C \sqsubseteq D$, *role inclusions* (RIs) of the form $r \sqsubseteq s$, and *assertions* of the form $A(a)$, $r(a,b)$, or $\rho(a,b)$. A *TBox* $\mathcal{T}$ is a finite set of GCIs and RIs, and an *ABox* $\mathcal{A}$ is a finite set of assertions. Together, they form a *knowledge base* (KB) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$.

Note that the indiscernibility relation ρ is not an element of the set of role names $\mathsf{N_R}$ and does not occur in TBoxes explicitly, but it can be used directly in ABoxes to state that

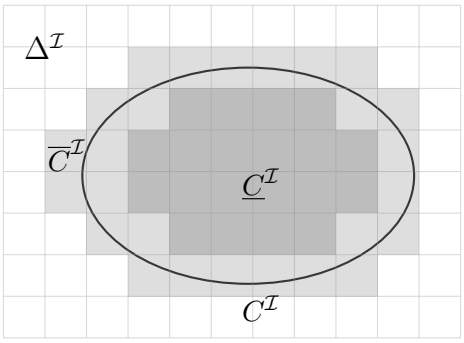

Figure 1: Semantics of a concept (ellipse), its upper (light grey) and lower (dark grey) approximation.

two objects cannot be distinguished. The relation ρ is the basis for the semantics of the upper and lower approximation.

We denote the sets of all concept names, role names, individual names, and concepts (including syntactic subconcepts) occurring in a set X of expressions by $\mathsf{N_C}(X)$, $\mathsf{N_R}(X)$, $\mathsf{N_I}(X)$, and $\mathbb{C}(X)$, respectively.

Semantics. An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty set $\Delta^{\mathcal{I}}$, called the *domain* of $\mathcal{I}$, and an *interpretation function* $\cdot^{\mathcal{I}}$, which 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 $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, to every $a \in \mathsf{N_I}$ an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ such that, for all $a, b \in \mathsf{N_I}$, $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ if $a \neq b$ (unique name assumption), and to ρ an equivalence relation $\rho^{\mathcal{I}}$ on $\Delta^{\mathcal{I}}$.

Let $[x]_\sim$ denote the *equivalence class* of $x \in \Delta^{\mathcal{I}}$ under the relation $\sim$. The function $\cdot^{\mathcal{I}}$ is extended to complex concepts by setting $\top^{\mathcal{I}} := \Delta^{\mathcal{I}}$, $\bot^{\mathcal{I}} := \emptyset$, and

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

The *granule* of an element $x \in \Delta^{\mathcal{I}}$ is the equivalence class $[x]_{\rho^{\mathcal{I}}}$ of elements indiscernible from x. Intuitively, $\overline{D}$ relaxes D to the union of all those granules with elements in D. Inversely, $\underline{D}$ strengthens $D^{\mathcal{I}}$ to those elements whose granule is fully contained in $D^{\mathcal{I}}$. Observe that the lower approximation behaves to some extent like a value restriction from more expressive DLs in the sense that it refers to *all elements of a granule*. The semantics of the upper approximation $\overline{C}$ and the lower approximation $\underline{C}$ are shown in Figure 1 in relation to concept C.

The interpretation $\mathcal{I}$ is a *model* of the GCI $C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, the RI $r \sqsubseteq s$ iff $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$, the assertion $A(a)$ iff $a^{\mathcal{I}} \in A^{\mathcal{I}}$ and the assertion $\hat{r}(a,b)$ with $\hat{r} \in \mathsf{N_R} \cup \{\rho\}$ iff $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in \hat{r}^{\mathcal{I}}$. An interpretation $\mathcal{I}$ is a *model* of (or *satisfies*) a set of axioms X, written $\mathcal{I} \models X$, iff it is a model of all axioms in X. A KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is *consistent* if $\mathcal{T} \cup \mathcal{A}$ has a model, and *inconsistent* otherwise. $\mathcal{K}$ *entails* an axiom α, written $\mathcal{K} \models \alpha$, iff all models of $\mathcal{K}$ also satisfy α. Given two concepts C and D, we say that C *subsumes* D w.r.t. $\mathcal{K}$ (written $C \sqsubseteq_{\mathcal{K}} D$), iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds in every model $\mathcal{I}$ of the KB $\mathcal{K}$.

Query Answering. Consider a set of *variables* $\mathsf{N_V}$ which is disjoint from $\mathsf{N_I} \cup \mathsf{N_C} \cup \mathsf{N_R}$, and let $\mathsf{N_T} := \mathsf{N_V} \cup \mathsf{N_I}$ be the set of *terms*. A *first-order (FO) query* is a FO formula $\Phi(\vec{x})$ over the signature $\mathsf{N_C} \cup \mathsf{N_R} \cup \{\rho\} \cup \mathsf{N_T}$.

The tuple $\vec{x} = x_1, \ldots, x_k$ with $x_i \in \mathsf{N_V}$ for all i, with $1 \leq i \leq k$ are the *answer variables* of $\Phi(\vec{x})$. A query containing k answer variables is a k-ary query. Let C be an $\mathcal{ELH}^{\rho}_{\bot}$ concept, $r \in \mathsf{N_R}$, and $t, u \in \mathsf{N_T}$. A *conjunctive query* (CQ) is a FO query of the form $\exists \vec{v}.\Phi(\vec{v}, \vec{w})$, where Φ is a (possibly empty) conjunction built of concept atoms $C(t)$, role atoms $r(t,u)$, and indiscernibility atoms $\rho(t,u)$. The empty conjunction is denoted by true.

Given an interpretation $\mathcal{I}$, a k-ary FO query $\Phi(\vec{x})$, and $a_i \in \mathsf{N_I}$ for i, with $1 \leq i \leq k$, we write $\mathcal{I} \models \Phi(a_1, \ldots, a_k)$ if the interpretation $\mathcal{I}$ satisfies $\Phi(\vec{x})$ with x_i assigned to $a_i^{\mathcal{I}}$ for i, with $1 \leq i \leq k$, and call $(a_1, \ldots, a_k)$ an *answer* to Φ in $\mathcal{I}$. Such a tuple $(a_1, \ldots, a_k)$ is a *certain answer* to Φ w.r.t. a KB $\mathcal{K}$ if, for every model $\mathcal{I}$ of $\mathcal{K}$, we have $\mathcal{I} \models \Phi(a_1, \ldots, a_k)$. The set $\mathsf{Cert}(\Phi, \mathcal{K})$ contains all certain answers for a given CQ Φ w.r.t. a KB $\mathcal{K}$.

The reasoning task investigated in this paper is *CQ answering* in $\mathcal{ELH}^{\rho}_{\bot}$, i.e., the computation of the set $\mathsf{Cert}(\Phi, \mathcal{K})$. In contrast to $\mathcal{EL}$ queries, the CQs considered here can contain indiscernibility atoms $\rho(t,u)$ and complex concepts using upper or lower approximation, provided they have a concept name assigned in the TBox.

When convenient, we view a CQ Φ as the set of atoms occurring in it. For a given query Φ we use the following sets: $\mathsf{N_T}(\Phi)$ for its terms, $\mathsf{N_V}(\Phi)$ for its variables, $\mathsf{N_{AV}}(\Phi)$ for its answer variables and $\mathsf{N_{QV}}(\Phi)$ for its quantified variables.

The Combined Approach for $\mathcal{ELH}^{\rho}_{\bot}$

Recall that the combined approach for query answering first absorbs the TBox information into the ABox. Afterwards, it computes a query rewriting that augments the initial query by filter conditions.

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a KB. For the remainder of the paper we make the following simplifying assumptions w.l.o.g.

1. CQs over $\mathcal{K}$ contain only individual names that do occur in $\mathcal{K}$,
2. $\mathcal{K}$ contains no role synonyms; i.e., there are no $r, s \in \mathsf{N_R}$ such that $r \neq s$ and $\mathcal{K} \models \{r \sqsubseteq s, s \sqsubseteq r\}$,[1] and
3. all concept names that appear in $\mathcal{A}$ appear also in $\mathcal{T}$.

For the rest of the paper let Φ be a k-ary CQ to be answered w.r.t. a consistent $\mathcal{ELH}^{\rho}_{\bot}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$.

Absorption of TBox Axioms

The goal of TBox absorption is to rewrite the ABox in such a way that all the background knowledge is already included in it. In this way, the TBox can be disregarded in the query answering process, using only the relevant information encoded in the rewritten ABox. We show how this method,

$$\Delta^{\mathcal{I}_\mathcal{K}} := \mathsf{N}_\mathsf{I}(\mathcal{A}) \cup \mathsf{N}_\mathsf{I}^{\mathbb{C}} \cup \mathsf{N}_\mathsf{I}^{\mathsf{low}} \cup \mathsf{N}_\mathsf{I}^{\mathsf{up}}$$

$$a^{\mathcal{I}_\mathcal{K}} := a$$

$$A^{\mathcal{I}_\mathcal{K}} := \{a \in \mathsf{N}_\mathsf{I}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup \{x_C \in \mathsf{N}_\mathsf{I}^{\mathbb{C}} \mid \mathcal{K} \models C \sqsubseteq A\} \cup \{x_{C,e} \in \mathsf{N}_\mathsf{I}^{\mathsf{up}} \mid \mathcal{K} \models C \sqsubseteq A\} \cup \{x_{C,b}, \ell_b \in \mathsf{N}_\mathsf{I}^{\rho} \mid \mathcal{K} \models \underline{A}(b)\} \cup \{x_{C,x_D}, \ell_{x_D} \in \mathsf{N}_\mathsf{I}^{\rho} \mid \mathcal{K} \models D \sqsubseteq \underline{A}\}$$

$$r^{\mathcal{I}_\mathcal{K}} := \{(a,b) \mid s(a,b) \in \mathcal{A},\ \mathcal{K} \models s \sqsubseteq r\} \cup \{(a, x_C) \in \mathsf{N}_\mathsf{I}(\mathcal{A}) \times \mathsf{N}_\mathsf{I}^{\mathbb{C}} \mid \mathcal{K} \models \exists r.C(a)\} \cup \{(x_C, x_D) \in \mathsf{N}_\mathsf{I}^{\mathbb{C}} \times \mathsf{N}_\mathsf{I}^{\mathbb{C}} \mid \mathcal{K} \models C \sqsubseteq \exists r.D\} \cup \{(x_{C,e}, x_D) \in \mathsf{N}_\mathsf{I}^{\mathsf{up}} \times \mathsf{N}_\mathsf{I}^{\mathbb{C}} \mid \mathcal{K} \models C \sqsubseteq \exists r.D\} \cup \{(x_{C,b}, x_D), (\ell_b, x_D) \in \mathsf{N}_\mathsf{I}^{\rho} \times \mathsf{N}_\mathsf{I}^{\mathbb{C}} \mid \mathcal{K} \models \underline{\exists r.\underline{D}}(b)\} \cup \{(x_{C,x_E}, x_D), (\ell_{x_E}, x_D) \in \mathsf{N}_\mathsf{I}^{\rho} \times \mathsf{N}_\mathsf{I}^{\mathbb{C}} \mid \mathcal{K} \models E \sqsubseteq \underline{\exists r.\underline{D}}\}$$

$$\rho_\mathcal{K} := \{(a,b) \mid \rho(a,b) \in \mathcal{A}\} \cup \{(a, x_{C,a}) \in \mathsf{N}_\mathsf{I}(\mathcal{A}) \times \mathsf{N}_\mathsf{I}^{\mathsf{up}} \mid \mathcal{K} \models \overline{C}(a)\} \cup \{(e, \ell_e) \mid \ell_e \in \mathsf{N}_\mathsf{I}^{\mathsf{low}}\} \cup \{(x_C, x_{D,x_C}) \in \mathsf{N}_\mathsf{I}^{\mathbb{C}} \times \mathsf{N}_\mathsf{I}^{\mathsf{up}} \mid \mathcal{K} \models C \sqsubseteq \overline{D}\} \cup \{(x_{C,e}, x_{D,e}) \in \mathsf{N}_\mathsf{I}^{\mathsf{up}} \times \mathsf{N}_\mathsf{I}^{\mathsf{up}} \mid \mathcal{K} \models C \sqsubseteq \overline{D}\}$$

$$\rho^{\mathcal{I}_\mathcal{K}} := \text{ reflexive, symmetric, transitive closure of } \rho_\mathcal{K}$$

Figure 2: The canonical interpretation $\mathcal{I}_\mathcal{K} = (\Delta^{\mathcal{I}_\mathcal{K}}, \cdot^{\mathcal{I}_\mathcal{K}})$ of $\mathcal{K}$, where $a \in \mathsf{N}_\mathsf{I}(\mathcal{A})$, $A \in \mathsf{N}_\mathsf{C}(\mathcal{K})$, and $r \in \mathsf{N}_\mathsf{R}(\mathcal{K})$.

originally devised for $\mathcal{EL}$, can be lifted to rough DL $\mathcal{ELH}^\rho_\bot$.

ABox rewritings are usually represented as canonical interpretations. The canonical interpretations (Lutz and Wolter 2010) used in the combined approach for $\mathcal{EL}$ (Lutz, Toman, and Wolter 2009), need to be extended for $\mathcal{ELH}^\rho_\bot$ to accommodate the information from the upper and lower approximation concept constructors and from the ρ-assertions in the ABox. Canonical models that treat upper and lower approximations were previously described in (Peñaloza and Zou 2013), where the goal was to decide concept subsumption and thus the focus was on the TBox only. For our case these canonical models need to be extended to represent the information from the (input) ABox, too.

To formally define the canonical interpretations, we must introduce the *normal form*. We say that a TBox is in normal form if all its GCIs are of the form

$$A \sqcap B \sqsubseteq C, \qquad \exists r.A \sqsubseteq B, \qquad A \sqsubseteq \exists r.B,$$
$$A \sqsubseteq \underline{B}, \qquad A \sqsubseteq \overline{B}, \qquad \underline{A} \sqsubseteq B,$$

where A, B are concept names or $\top$ and C is a concept name, $\bot$ or $\top$. Every $\mathcal{ELH}^\rho_\bot$ TBox can be transformed to normal form in polynomial time. In the following we assume that the TBox is always in normal form.

The canonical interpretations of $\mathcal{ELH}^\rho_\bot$ contain four sorts of domain elements. We first give an overview of the sorts and then define the sets containing them. Two sorts are as in canonical interpretations for classical $\mathcal{EL}$: representatives for individual names occurring in the ABox $\mathcal{A}$ (collected in the set $\mathsf{N}_\mathsf{I}(\mathcal{A})$), and for concepts occurring in the TBox $\mathcal{T}$ (collected in $\mathsf{N}_\mathsf{I}^{\mathbb{C}}$). We call these elements *seed elements*. Additionally, we use two new sorts of domain elements: representatives for the lower approximations of each concept or individual occurring in the KB (collected in $\mathsf{N}_\mathsf{I}^{\mathsf{low}}$) and representatives for members of the upper approximations of concepts (collected in $\mathsf{N}_\mathsf{I}^{\mathsf{up}}$).

We turn now to the definition of the sets capturing these four sorts of domain elements. For simplicity, the *named elements* representing the individual names are denoted by the corresponding names from $\mathsf{N}_\mathsf{I}(\mathcal{A})$. The other elements are called *auxiliary elements* and are contained in the sets:

$$\mathsf{N}_\mathsf{I}^{\mathbb{C}} := \{x_C \mid C \in \mathbb{C}(\mathcal{T})\}$$
$$\mathsf{N}_\mathsf{I}^{\mathsf{up}} := \{x_{C,e} \mid C \in \mathbb{C}(\mathcal{T}), e \in \mathsf{N}_\mathsf{I}(\mathcal{A}) \cup \mathsf{N}_\mathsf{I}^{\mathbb{C}}\}$$
$$\mathsf{N}_\mathsf{I}^{\mathsf{low}} := \{\ell_e \mid e \in \mathsf{N}_\mathsf{I}(\mathcal{A}) \cup \mathsf{N}_\mathsf{I}^{\mathbb{C}}\}$$

Intuitively, the auxiliary elements stand for the following:

- $x_C \in \mathsf{N}_\mathsf{I}^{\mathbb{C}}$ represents an element that satisfies C and acts as role-successor; it is employed to make the predecessors satisfy concepts of the form $\exists r.C$;
- $x_{C,e} \in \mathsf{N}_\mathsf{I}^{\mathsf{up}}$ represents an element that satisfies C. If the seed element e is an individual, then $x_{C,e}$ is indiscernible from e. In the case where the seed element e is a concept, then $x_{C,e}$ represents that every element from e is indistinguishable from some element in C. The element $x_{C,e}$ is used to make the seed element e satisfy $\overline{C}$; and
- $\ell_e \in \mathsf{N}_\mathsf{I}^{\mathsf{low}}$ represents an element satisfying exactly those concepts C that are satisfied by all elements in the lower approximation of e. If seed element e is an individual, then ℓ_e is indiscernible from element e. If seed element e is a concept, then ℓ_e represents all granules fully contained in e. The seed element e satisfies $\underline{C}$ for all concepts C associated to ℓ_e.

Sometimes we use the short-hand $\mathsf{N}_\mathsf{I}^{\rho} = \mathsf{N}_\mathsf{I}^{\mathsf{up}} \cup \mathsf{N}_\mathsf{I}^{\mathsf{low}}$ for the 'non-seed' elements. Observe that all elements in $\mathsf{N}_\mathsf{I}^{\mathsf{up}}$ or $\mathsf{N}_\mathsf{I}^{\mathsf{low}}$ are 'caused' by a seed element. The idea is that in the canonical interpretation each seed element is associated with an element representing this seed element's lower approximation. ABox individuals have the same granule as their

[1]This can always be achieved by replacing in the KB $\mathcal{K}$ all occurrences of synonyms of role r by r.

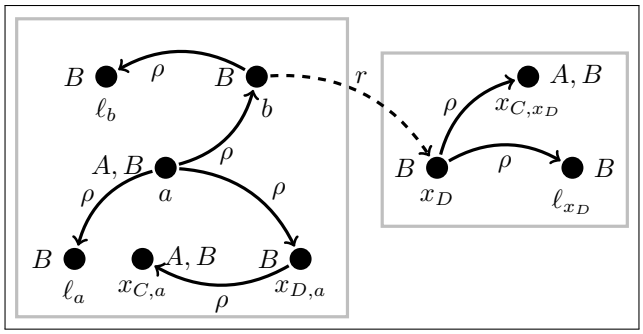

Figure 3: The canonical interpretation $\mathcal{I}_{\mathcal{K}_1}$ (without transitive, reflexive $\rho^{\mathcal{I}_{\mathcal{K}_1}}$-edges and C, D labels) for the KB $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A}_1)$ from Example 1 shown as a graph. Nodes represent domain elements and are labeled by the concept names they instantiate, edges represent relations. The gray frames highlight the granules of a and x_D.

lower or upper approximation, thus they only induce one element in $\mathsf{N}_\mathsf{I}^{\mathsf{low}}$. In contrast to this, concepts from $\mathcal{T}$ can have several granules in their approximations. Here, the lower approximation captures what is common to all granules in the lower approximation, thus one element in $\mathsf{N}_\mathsf{I}^{\mathsf{low}}$ representing the lower approximation of a concept suffices. The granules in the upper approximation of a concept C can overlap with different concepts or individuals e, thus different representatives for each such overlap are introduced in $\mathsf{N}_\mathsf{I}^{\mathsf{up}}$. During the reasoning process it can be discovered that some of the granule representatives belong into the same granule, which then gives rise to ρ-edges between the granule representatives.

The *canonical interpretation* $\mathcal{I}_\mathcal{K}$ of a KB $\mathcal{K}$ is formally defined in Figure 2, through a description of the interpretation function of all the relevant elements. The size of $\Delta^{\mathcal{I}_\mathcal{K}}$ is polynomial (more precisely, cubic) in the size of $\mathcal{K}$. Moreover, $\mathcal{I}_\mathcal{K}$ is computable in polynomial time, and consistency of $\mathcal{K}$ can be checked in polynomial time (Peñaloza and Zou 2013).

Example 1. *Consider* $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A}_1)$ *with* $\mathcal{T}_1 = \{D \sqsubseteq \overline{C}, C \sqsubseteq A \sqcap \underline{B}\}$, *and* $\mathcal{A}_1 = \{C(a), \overline{D}(a), \exists r.D(b), \rho(a,b)\}$.

Figure 3 depicts its canonical interpretation $\mathcal{I}_{\mathcal{K}_1}$ *(omitting transitive, reflexive* ρ*-edges). As in classical canonical interpretations,* a *is an instance of* A *since* $(*)$ $\mathcal{I}_{\mathcal{K}_1}$ *satisfies both* $C(a)$ *and* $C \sqsubseteq A \sqcap \underline{B}$. *E The element* b *is an instance of* $\exists r.D$, *since it is related to the representative instance of* D *(*$x_D \in \mathsf{N}_\mathsf{I}^{\mathbb{C}}$*) via* $r^{\mathcal{I}_{\mathcal{K}_1}}$.

In the rough setting, the relation $\rho^{\mathcal{I}_{\mathcal{K}_1}}$ *comes into play and* $(*)$ *yields that* a *is an instance of* $\underline{B}$*; i.e., all elements in* $[a]_\rho$, *especially* $\ell_a \in \mathsf{N}_\mathsf{I}^{\mathsf{low}}$, *instantiate* B *in* $\mathcal{I}_{\mathcal{K}_1}$. *Since* $D \sqsubseteq \overline{C}$, x_D *instantiates* $\overline{C}$*; i.e., it is related via* $\rho^{\mathcal{I}_{\mathcal{K}_1}}$ *to its representative* ρ*-successor instantiating* C*:* x_{C,x_D}. *The latter similarly holds for* $x_{D,a}$, *the representative* ρ*-successor of* a *that instantiates* D*; i.e.,* $x_{D,a}$ *is related via* $\rho^{\mathcal{I}_{\mathcal{K}_1}}$ *to its representative successor that is an instance of* C, $x_{C,a}$. *Note, that* $x_{D,a}$ *exists due to the assertion* $\overline{D}(a)$.

Figure 3 shows that canonical interpretations in $\mathcal{ELH}^\rho_\bot$ correspond to the ones in $\mathcal{EL}$ modulo the granules—by regarding each granule as a single element, the result is an $\mathcal{EL}$ interpretation that satisfies the TBox without approximation constructors. However, role assertions from the ABox can establish role edges between members of the same granule.

Lemma 2. *If* $\mathcal{K}$ *is consistent, then* $\mathcal{I}_\mathcal{K}$ *is a model of* $\mathcal{K}$.

Proof (Sketch). By construction, $\mathcal{I}_\mathcal{K}$ is a model of $\mathcal{A}$ and of all RIs in $\mathcal{T}$. We need to show that the GCIs in $\mathcal{T}$ are satisfied. By induction on the concept structure it can be shown that, for all $C \in \mathbb{C}(\mathcal{T})$, $a \in \mathsf{N}_\mathsf{I}(\mathcal{A})$, and $x_E \in \mathsf{N}_\mathsf{I}^{\mathbb{C}}$, it holds that $a \in C^{\mathcal{I}_\mathcal{K}}$ iff $\mathcal{K} \models C(a)$, and $x_E \in C^{\mathcal{I}_\mathcal{K}}$ iff $\mathcal{K} \models E \sqsubseteq C$. Similar equivalences hold for elements of the form $x_{C,e}$ and ℓ_e. Then, it is easy to show that the GCIs $C \sqsubseteq D$ are satisfied by applying the corresponding equivalences to C and D. □

As mentioned already, the interpretation $\mathcal{I}_\mathcal{K}$ can be seen as an ABox that encodes all the information stated in the original KB $\mathcal{K}$. However, queries cannot be answered using $\mathcal{I}_\mathcal{K}$ directly, for two reasons. The first reason is, that the domain $\Delta^{\mathcal{I}_\mathcal{K}}$ of $\mathcal{I}_\mathcal{K}$ may contain superfluous elements. For example, for the KB $\mathcal{K}_2 = (\{C \sqsubseteq A\}, \emptyset)$, $\mathcal{I}_{\mathcal{K}_2}$ contains an element $x_C \in A^{\mathcal{I}_{\mathcal{K}_2}}$. Thus, the CQ $\Phi_2 = \exists y.A(y)$ would return an empty tuple (meaning that the query can be satisfied) w.r.t. $\mathcal{I}_{\mathcal{K}_2}$, even though this is not an answer w.r.t. $\mathcal{K}_2$. We therefore restrict the canonical model $\mathcal{I}_\mathcal{K}$ to those domain elements that are reachable from named elements.

A *path* in an interpretation $\mathcal{I}$ is a finite sequence $d_0 \widehat{r_1} d_1 \cdots \widehat{r_n} d_n$, $n \geq 0$, such that $d_0 \in \mathsf{N}_\mathsf{I}(\mathcal{A})$ and, for all i with $1 \leq i \leq n$, $d_i \in \Delta^\mathcal{I} \setminus \mathsf{N}_\mathsf{I}(\mathcal{A})$, $\widehat{r_i} \in \mathsf{N}_\mathsf{R} \cup \{\rho_\mathcal{K}\}$, and $(d_{i-1}, d_i) \in \widehat{r_i}^{\mathcal{I}}$. $\mathsf{Paths}(\mathcal{I})$ denotes the set of all paths in $\mathcal{I}$. For a path $p = d_0 r_1 d_1 \cdots r_n d_n$, define $\mathsf{Tail}(p) := d_n$. Intuitively, each path starts with an element that represents an ABox individual, each such element starts a path and there is no second ABox individual on a path. Observe that paths are defined using $\rho_\mathcal{K}$ and not its symmetric, reflexive, transitive closure.

To avoid the superfluous domain elements, the interpretation $\mathcal{I}^{\mathsf{re}}_\mathcal{K}$ is obtained from $\mathcal{I}_\mathcal{K}$ by restricting its domain elements to those reachable from elements that represent ABox individuals, or, more formally:

$$\Delta^{\mathcal{I}^{\mathsf{re}}_\mathcal{K}} = \{\mathsf{Tail}(p) \mid p \in \mathsf{Paths}(\mathcal{I}_\mathcal{K})\}.$$

The next fact follows directly from this definition and states for those seed elements reachable by paths, the members of their granule. Thus it clarifies the picture of the indiscernibility relation in $\mathcal{I}^{\mathsf{re}}_\mathcal{K}$.

Fact 3. *For all seed elements that are reachable by paths, i.e., for all* $e \in \mathsf{N}_\mathsf{I}(\mathcal{A}) \cup (\mathsf{N}_\mathsf{I}^{\mathbb{C}} \cap \Delta^{\mathcal{I}^{\mathsf{re}}_\mathcal{K}})$, *we have*

$$[e]_{\rho^{\mathcal{I}^{\mathsf{re}}_\mathcal{K}}} = \{e, \ell_e\} \cup \{x_{D,e} \in \mathsf{N}_\mathsf{I}^{\mathsf{up}} \cap \Delta^{\mathcal{I}^{\mathsf{re}}_\mathcal{K}}\} \cup \bigcup_{\rho(e,a) \in \mathcal{A}} \Big(\{a, \ell_a\} \cup \{x_{D,a} \in \mathsf{N}_\mathsf{I}^{\mathsf{up}} \cap \Delta^{\mathcal{I}^{\mathsf{re}}_\mathcal{K}}\}\Big).$$ □

The second reason why queries cannot be answered using $\mathcal{I}_\mathcal{K}$ directly, is the unintended reuse of some elements. In the classical case of $\mathcal{EL}$, the elements in $\mathsf{N}_\mathsf{I}^{\mathbb{C}}$ can introduce unintended joins in the model, and hence yield erroneous

answers. As noticed in (Lutz, Toman, and Wolter 2009), for the KB $\mathcal{K}_3 = (\mathcal{T}_3, \mathcal{A}_3)$ with $\mathcal{T}_3 = \{A \sqsubseteq \exists r.B \sqcap \exists s.B\}$ and $\mathcal{A}_3 = \{A(a)\}$, the element a is connected to x_B via r and s in $\mathcal{I}_{\mathcal{K}_3}$. Considering the query $\Phi_3(x) = \exists y.r(x,y) \wedge s(x,y)$, this gives rise to $\mathcal{I}_{\mathcal{K}_3} \models \Phi_3(a)$, but $a \notin \mathsf{Cert}(\Phi_3, \mathcal{K}_3)$.

In the $\mathcal{ELH}^{\rho}_{\bot}$ case, with the interpretation $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$, the unintended reuse additionally affects those elements from $\mathsf{N}^{\rho}_{\mathsf{I}}$ (connected to the $\mathsf{N}^{\mathbb{C}}_{\mathsf{I}}$-elements) that were induced by seed elements from $\mathsf{N}^{\mathbb{C}}_{\mathsf{I}}$. So, for the KB $\mathcal{K}_3$, there would be (among others) the element $x_{B,x_B} \in \mathsf{N}^{\mathsf{up}}_{\mathsf{I}}$ in the domain of $\mathcal{I}_{\mathcal{K}_3}$. This element is connected to element x_B by a ρ-edge. For the query $\Phi'_3(x) = \exists y.r(x,y) \wedge s(x,y) \wedge \overline{B}(y)$ this gives rise to $\mathcal{I}_{\mathcal{K}_3} \models \Phi'_3(a)$, but $a \notin \mathsf{Cert}(\Phi'_3, \mathcal{K}_3)$.

To remedy these effects, the canonical model can be *unraveled* into a new, tree-shaped interpretation $\mathcal{U}_{\mathcal{K}}$ so that the paths in $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$ become the domain elements of $\mathcal{U}_{\mathcal{K}}$. The *unraveling* of $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$ is the interpretation $\mathcal{U}_{\mathcal{K}} = (\Delta^{\mathcal{U}_{\mathcal{K}}}, \cdot^{\mathcal{U}_{\mathcal{K}}})$, where, for all $a \in \mathsf{N}_{\mathsf{I}}(\mathcal{A}), A \in \mathsf{N}_{\mathsf{C}}(\mathcal{K}), r \in \mathsf{N}_{\mathsf{R}}(\mathcal{K})$:

$$
\begin{aligned}
\Delta^{\mathcal{U}_{\mathcal{K}}} &:= \mathsf{Paths}(\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}) \\
a^{\mathcal{U}_{\mathcal{K}}} &:= a \\
A^{\mathcal{U}_{\mathcal{K}}} &:= \{p \mid \mathsf{Tail}(p) \in A^{\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}}\} \\
r^{\mathcal{U}_{\mathcal{K}}} &:= \{(a,b) \mid a,b \in \mathsf{N}_{\mathsf{I}}(\mathcal{A}), (a,b) \in r^{\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}}\} \cup \\
&\quad \{(p, p \cdot se) \mid p, p \cdot se \in \Delta^{\mathcal{U}_{\mathcal{K}}}, \mathcal{K} \models s \sqsubseteq r\} \\
\rho_{\mathcal{K}'} &:= \{(a,b) \mid a,b \in \mathsf{N}_{\mathsf{I}}(\mathcal{A}), (a,b) \in \rho^{\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}}\} \cup \\
&\quad \{(p, p \cdot \rho e) \mid p \cdot \rho e \in \Delta^{\mathcal{U}_{\mathcal{K}}}\} \\
\rho^{\mathcal{U}_{\mathcal{K}}} &:= \text{ reflexive, symmetric, transitive closure of } \rho_{\mathcal{K}'}
\end{aligned}
$$

Note that the construction of $\mathcal{U}_{\mathcal{K}}$ from $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$ does not depend on the GCIs but only on the RIs in $\mathcal{T}$.

Lemma 4. *For every $a_1, \ldots, a_k \in \mathsf{N}_{\mathsf{I}}(\mathcal{A})$, we have that $(a_1, \ldots, a_k) \in \mathsf{Cert}(\Phi, \mathcal{K})$ iff $\mathcal{U}_{\mathcal{K}} \models \Phi[a_1, \ldots, a_k]$.*

The unraveling $\mathcal{U}_{\mathcal{K}}$ gives the correct answers to CQs, but it is typically infinite; e.g. in the presence of terminological cycles. The idea is therefore to focus on $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$ for CQ answering, but to take $\mathcal{U}_{\mathcal{K}}$ as a kind of reference model. Specifically, the query Φ is extended with conditions that accept only answers compliant with $\mathcal{U}_{\mathcal{K}}$, by avoiding the unintended joins.

The Query Rewriting

We focus now on the problem of *rewriting* a CQ Φ in such a way that the answers of its rewriting $\Phi^{\dagger}_{\mathcal{R}}$ w.r.t. $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$ correspond exactly to the answers of the original query Φ w.r.t. $\mathcal{K}$. More precisely, we want to prove the following result.

Theorem 5. *For every finite set of role inclusions $\mathcal{R}$ and each k-ary CQ Φ, one can construct in polynomial time a k-ary FO query $\Phi^{\dagger}_{\mathcal{R}}$ such that, for all $\mathcal{ELH}^{\rho}_{\bot}$ KBs $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ using only the role inclusions $\mathcal{R}$, and all $a_1, \ldots, a_k \in \mathsf{N}_{\mathsf{I}}(\mathcal{A})$, we have*

$$(a_1, \ldots, a_k) \in \mathsf{Cert}(\Phi, \mathcal{K}) \text{ iff } \mathcal{I}^{\mathsf{re}}_{\mathcal{K}} \models \Phi^{\dagger}_{\mathcal{R}}(a_1, \ldots, a_k).$$

In order to show this theorem, our first step is to develop the rewriting procedure. The combined approach extends a given CQ with additional filter conditions to discard those answers to Φ in $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$ that are not answers in $\mathcal{U}_{\mathcal{K}}$. These conditions essentially target those parts of the CQ that can be satisfied by non-tree structures that may exist in $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$ but not in $\mathcal{U}_{\mathcal{K}}$. Observe that only non-tree structures including auxiliary elements are critical as these are the ones that would not appear in the original KB. We extend the filter conditions from Lutz, Toman, and Wolter to handle also the elements representing upper and lower approximations of concepts.

Specifically, due to the properties of the indiscernibility relation ρ (i.e., transitivity, reflexivity, and symmetry), and its influence in the approximation constructors, the new filter conditions need to consider potential equivalences and joins that are only implicitly stated. For instance, a tree shaped query that leads to two different but indiscernible elements will include an implicit join that must be taken into account.

Let $\mathcal{R}$ be an arbitrary but fixed finite set of RIs and Φ be a k-ary CQ. To identify auxiliary elements, we introduce two fresh unary predicates (that is, concepts): Aux identifies elements from $\mathsf{N}^{\mathbb{C}}_{\mathsf{I}}$ and Aux_{ρ} 'approximation-related', i.e., 'non-seed' elements from $\mathsf{N}^{\rho}_{\mathsf{I}}$. We define them to be interpreted in $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$ and $\mathcal{U}_{\mathcal{K}}$ as:

$$
\begin{aligned}
\mathsf{Aux}^{\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}} &:= \Delta^{\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}} \cap \mathsf{N}^{\mathbb{C}}_{\mathsf{I}} \\
\mathsf{Aux}_{\rho}{}^{\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}} &:= \Delta^{\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}} \cap \mathsf{N}^{\rho}_{\mathsf{I}} \\
\mathsf{Aux}^{\mathcal{U}_{\mathcal{K}}} &:= \bigcup_{p \in \Delta^{\mathcal{U}_{\mathcal{K}}}, \mathsf{Tail}(p) \in \mathsf{N}^{\mathbb{C}}_{\mathsf{I}}} \{p\} \\
\mathsf{Aux}_{\rho}{}^{\mathcal{U}_{\mathcal{K}}} &:= \bigcup_{p \in \Delta^{\mathcal{U}_{\mathcal{K}}}, \mathsf{Tail}(p) \in \mathsf{N}^{\rho}_{\mathsf{I}}} \{p\}
\end{aligned}
$$

To model the filters, we describe those mappings from answer variables to ABox individuals that describe non-tree structures which cannot be satisfied in $\mathcal{U}_{\mathcal{K}}$. The latter is the case if the answer mapping uses a single $\mathsf{N}^{\mathbb{C}}_{\mathsf{I}}$ element as a role successor for mapping several objects referred to in the query such that there is no corresponding element in $\mathcal{U}_{\mathcal{K}}$ that fits all of them. A corresponding such element in $\mathcal{U}_{\mathcal{K}}$ exists, if the structures from the query can be mapped into a single path in $\mathcal{I}^{\mathsf{re}}_{\mathcal{K}}$, by identifying terms.

The terms that are identified in this way, and those that are indiscernible, are captured via an equivalence relation $\sim^{r}_{\Phi}$ on terms, grouping them into equivalence classes. Let $\sim^{\rho}_{\Phi}$ be another equivalence relation over $\mathsf{N}_{\mathsf{T}}(\Phi)$ induced by the atoms of the form $\rho(s,t)$ occurring in Φ for some terms s and t. The relation $\sim^{r}_{\Phi}$ is defined inductively based on $\sim^{\rho}_{\Phi}$ as the smallest transitive and reflexive relation on $\mathsf{N}_{\mathsf{T}}(\Phi)$ that (1) includes the relation

$$\{(t,t') \mid r_1(s,t), r_2(s',t') \in \Phi, r_1, r_2 \in \mathsf{N}_{\mathsf{R}}, t \sim^{\rho}_{\Phi} t'\}$$

and (2) satisfies the closure condition:

$$\text{if } r_1(s,t), r_2(s',t') \in \Phi, r_1, r_2 \in \mathsf{N}_{\mathsf{R}} \text{ and } t \sim^{r}_{\Phi} t', \text{ then } s \sim^{r}_{\Phi} s'. \qquad (\dagger)$$

Observe that the relation $\sim^{r}_{\Phi}$ inherits symmetry by construction from the symmetric relation $\sim^{\rho}_{\Phi}$ and, furthermore, $\sim^{r}_{\Phi}$ does not need to contain $\sim^{\rho}_{\Phi}$ as a sub-relation. The equivalence classes of $\sim^{r}_{\Phi}$ group those terms that cannot be distinguished by homomorphisms from Φ into $\mathcal{U}_{\mathcal{K}}$. Such an inductively defined relation is already used in the combined

approach for $\mathcal{EL}$ (Lutz, Toman, and Wolter 2009). The important difference is that in that previous work, the induction is based on the identity relation. The closure condition then captures non-tree structures in the query Φ, where a term t has two role-predecessors s and s'. For $\mathcal{ELH}^{\rho}_{\bot}$, the identity relation is too fine-grained, since truly distinct objects belong to different granules. So, in order to be able to handle in the query the relaxation introduced by the rough constructors, we need to consider the whole indiscernibility relation on the query terms. Since granules can be separated by role relationships (as shown in Figure 3), the incoming role edges of a granule and the related role-predecessor need to be addressed. In order to do so we define for each equivalence class ζ of the relation $\sim^r_\Phi$ the predicates:

$$\mathsf{Pre}(\zeta) := \{t \mid r(t,t') \in \Phi, r \in \mathsf{N_R}, t' \in \zeta\}$$
$$\mathsf{In}(\zeta) := \{r \mid r(t,t') \in \Phi, r \in \mathsf{N_R}, t' \in \zeta\}$$

The set $\mathsf{Pre}(\zeta)$ describes all the role predecessors of terms in the equivalence class ζ. The set $\mathsf{In}(\zeta)$ contains all the incoming role names to ζ.

For the roles that separate the granules, the role hierarchy $\mathcal{R}$ needs to be taken into account. As the more general role relationships of another is directly stated in the canonical model (by construction of $r^{\mathcal{I}_\mathcal{K}}$) and thus also in $\mathcal{I}^{\mathsf{re}}_\mathcal{K}$, the query needs to refer a most specific role. A role $r \in \mathsf{N_R}$ is an *implicant* of $R \subseteq \mathsf{N_R}$ if $\mathcal{R} \models r \sqsubseteq s$ for all $s \in R$. It is a *prime implicant* if, additionally, $\mathcal{R} \not\models r \sqsubseteq r'$ for all implicants r' of R with $r \neq r'$. Since KBs contains no role synonyms, there is a prime implicant for each $R \subseteq \mathsf{N_R}$ for which there is an implicant.

The different filters focus on different kinds of structures in Φ. We collect these structures in the following sets, which are based on the sets $\mathsf{Pre}(\zeta)$ and $\mathsf{In}(\zeta)$, and on implicants:

- $\mathsf{Fork}_{\neq}$ is the set of variables $v \in \mathsf{N_{QV}}(\Phi)$ such that there is no implicant of $\mathsf{In}([v]_{\sim^r_\Phi})$. Intuitively, $\mathsf{Fork}_{\neq}$ collects those variables that can never be mapped to the same Aux-element in $\mathcal{U}_\mathcal{K}$, due to the shape of Φ (i.e., there are different role atoms where the variables occur as successors) and the interpretation of roles in $\mathcal{U}_\mathcal{K}$, which is based on the RIs entailed by $\mathcal{K}$.
- $\mathsf{Fork}_{=}$ is the set of pairs $(\mathsf{Pre}(\zeta), \zeta)$ with $|\mathsf{Pre}(\zeta)| \geq 2$. The first terms in the pairs in $\mathsf{Fork}_{=}$ are those variables that are mapped to indiscernible elements by any homomorphism of Φ into $\mathcal{U}_\mathcal{K}$ and that may have to be identified if the successor variable is mapped to an Aux-element. Note that the case where the latter is not possible is captured by $\mathsf{Fork}_{\neq}$. Moreover, it does not suffice to require the identification, this is addressed next.
- $\mathsf{Fork}_{\mathcal{H}}$ 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 $\mathsf{In}(\zeta)$. By the definition of $\mathcal{U}_\mathcal{K}$, a pair of an arbitrary element and an element of $\mathsf{N}^{\mathbb{C}}_{\mathsf{I}}$ can be contained in the interpretations of different roles in $\mathcal{U}_\mathcal{K}$, but then it must also be in the interpretation of a prime implicant of those roles. $\mathsf{Fork}_{\mathcal{H}}$ therefore collects all relevant prime implicants so that the filter can enforce some such relation.
- Cyc is the set of all those quantified variables $v \in \mathsf{N_{QV}}(\Phi)$ such that there exist the role atoms $r_0(t_0, t'_0), \ldots, r_m(t_m, t'_m), \ldots, r_n(t_n, t'_n)$, $m, n \geq 0$ in Φ with $r_i \in \mathsf{N_R}$ for all $i, 0 \leq i \leq m$, and the following conditions hold:
  1. $(v, t_i) \in \sim^r_\Phi \cup \sim^\rho_\Phi$ for some $i \leq n$,
  2. $(t'_i, t_{i+1}) \in \sim^r_\Phi \cup \sim^\rho_\Phi$ for all $i < n$, and
  3. $(t'_n, t_m) \in \sim^r_\Phi \cup \sim^\rho_\Phi$;

  i.e., Cyc is the set of all quantified variables appearing in the query Φ that lead, through role connections and equivalences based on the indiscernibility relation, to cyclic dependencies.

These definitions are analogous to those employed in the combined approach for $\mathcal{EL}$; the main change in our setting is the integration of the indiscernibility relation into $\sim^r_\Phi$ to capture the notion of granules, which is fundamental for the correctness of the method. Notice that dealing with the indiscernibility relation requires a non-trivial extension of the classical case; indeed, indiscernible elements may affect many different points in the rewriting of a query. Moreover, to keep the connection to the work by Lutz, Toman, and Wolter explicit, we have used the same names for the filters; but they all differ from the original definitions.

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

Using these filters, we can now describe the promised query rewriting. Given the CQ Φ, we define the FO query

$$\Phi^\dagger_\mathcal{R} := \exists \vec{x}.(\Phi' \wedge \Psi_1 \wedge \Psi_2 \wedge \Psi_3), \text{ where}$$

$$\Psi_1 := \bigwedge_{v \in \mathsf{N_{AV}}(\Phi) \cup \mathsf{Fork}_{\neq} \cup \mathsf{Cyc}} \neg \mathsf{Aux}(v) \wedge \bigwedge_{v \in \mathsf{N_{AV}}(\Phi)} \neg \mathsf{Aux}_\rho(v)$$

$$\Psi_2 := \bigwedge_{(\{t_1, \ldots t_k\}, \zeta) \in \mathsf{Fork}_{=}} (\mathsf{Aux}(t_\zeta) \rightarrow \bigwedge_{1 \leq i < k} t_i = t_{i+1})$$

$$\Psi_3 := \bigwedge_{(\mathsf{I}, \zeta) \in \mathsf{Fork}_{\mathcal{H}}} (\mathsf{Aux}(t_\zeta) \rightarrow \bigvee_{r \in \mathsf{I}} r(t^{\mathsf{Pre}}_\zeta, t_\zeta)),$$

and Φ' is a CQ equivalent to Φ whose concept atoms are of the form $A(t)$ with $A \in \mathsf{N_C}$. This query Φ' it can be obtained from Φ through an unfolding that transforms complex concepts into first-order terms. For example, the unfolding rewrites the conjunct $\overline{C}(x)$ in Φ into $\exists y.\rho(x,y) \wedge C(y)$. Notice that the constraints enforcing that the explicit indiscernibility relations included in the original KB form an equivalence relation are already encoded in the definition of $\mathsf{Aux}_\rho{}^{\mathcal{I}^{\mathsf{re}}_\mathcal{K}}$ and $\mathsf{Aux}_\rho{}^{\mathcal{U}_\mathcal{K}}$.

The proof of Theorem 5 focuses on the new query $\Phi^\dagger_\mathcal{R}$, which can, in fact, be constructed in polynomial time. It remains to show that this query satisfies the property claimed by the theorem. The idea is that the filter conditions introduced in the rewriting make sure that the answers over Φ that do not hold in $\mathcal{U}_\mathcal{K}$ are excluded. Ψ_1 sifts out those answers in $\mathcal{I}^{\mathsf{re}}_\mathcal{K}$ that contain auxiliary elements, and those that cannot be mirrored in $\mathcal{U}_\mathcal{K}$ because the corresponding mapping uses an $\mathsf{N}^{\mathbb{C}}_{\mathsf{I}}$ element as a role successor in several cases such

that there is no corresponding element in $\mathcal{U}_\mathcal{K}$ that fits all of them. The query parts Ψ_2 and Ψ_3 characterize the situation in which a corresponding element in $\mathcal{U}_\mathcal{K}$ exists: by identifying elements, the relevant structures from Φ' mapped into $\mathcal{I}^{\mathsf{re}}_\mathcal{K}$ must be collapsible into a single path (Ψ_2), and a prime implicant must be among the edges between two nodes of this path (Ψ_3).

The proof of Theorem 5 uses the FO query $\Phi^\dagger_\mathcal{R}$. The filtering conditions introduced in this rewriting make sure that the answers over Φ that do not hold in the model $\mathcal{U}_\mathcal{K}$ are excluded. For example, Ψ_1 guarantees, amongst others, that any cyclic dependency between domain elements must occur in the ABox. That is, cycles introduced by the reuse of auxiliary names in the canonical model are ignored. For lack of space, full proofs are not included in this paper. They can be found in the appendix uploaded with this submission.

We now provide some simple examples of the rewriting, aimed to explain the ideas of the construction. Let $\mathcal{R} = \emptyset$. Notice that in this case, $\mathsf{Fork}_\mathcal{H}$ is always empty, and hence $\Psi_3 = \mathsf{true}$. We omit $\mathcal{R}$ and these Ψ_3 formulas in the rewritings. We first demonstrate the role of Cyc. Consider

$$\Phi_4 := \exists y_1, y_2.(\mathsf{hasA}(y_1, y_2) \land \rho(y_1, y_2)).$$

We have $\mathsf{Cyc} = \{y_1, y_2\}$, $\mathsf{Fork}_= = \mathsf{Fork}_\neq = \mathsf{Fork}_\mathcal{H} = \emptyset$, and thus obtain the following rewriting $\Phi^\dagger_4$:

$$\exists y_1, y_2.(\mathsf{hasA}(y_1, y_2) \land \rho(y_1, y_2) \land \neg\mathsf{Aux}(y_1) \land \neg\mathsf{Aux}(x_2)).$$

This query guarantees that all the answer pairs provided are indiscernible elements, related via the role hasA, and that they contain no auxiliary elements. We next consider a similar query, demonstrating the rewriting of forking situations:

$$\Phi_5 := \exists y_1, y_2.(\mathsf{hasA}(x_1, y_1) \land \mathsf{hasA}(x_2, y_2) \land \rho(y_1, y_2)).$$

The relation $\sim^\rho_{\Phi_5}$ has equivalence classes $\{x_1\}$, $\{x_2\}$, and $\{y_1, y_2\}$, and $\sim^r_{\Phi_5}$ defines the partition $\{\{x_1, x_2\}, \{y_1, y_2\}\}$. $\mathsf{Pre}(\{y_1, y_2\}) = \{x_1, x_2\}$ and $\mathsf{In}(\{y_1, y_2\}) = \{\mathsf{hasA}\}$. Thus, we have $\mathsf{Fork}_= = \{(\{x_1, x_2\}, \{y_1, y_2\})\}$, and $\mathsf{Fork}_\neq = \mathsf{Fork}_\mathcal{H} = \mathsf{Cyc} = \emptyset$. This yields the rewriting

$$\begin{aligned}\Phi^\dagger_5 =& \exists y_1, y_2.(\mathsf{hasA}(x_1, y_1) \land \mathsf{hasA}(x_2, y_2) \land \rho(y_1, y_2) \land \\ & \neg\mathsf{Aux}(x_1) \land \neg\mathsf{Aux}(x_1) \land (\mathsf{Aux}(y_1) \to x_1 = x_2)).\end{aligned}$$

Notice that every step in the construction of the rewriting is polynomial in the size of the KB and the query. Specifically, $\sim^r_\Phi$, Pre, and In are subsets of terms and variables that appear explicitly in Φ. By extension, the filters $\mathsf{Fork}_\neq$, $\mathsf{Fork}_=$, $\mathsf{Fork}_\mathcal{H}$, and Cyc are also polynomial in Φ. The only remaining case is ensuring that the auxiliary elements are not used to generate non-existing answers, as guaranteed by the queries $\Psi_i, 1 \leq i \leq 3$. The size of these queries is, in fact, polynomial in the number of auxiliary variables in $\mathcal{I}^{\mathsf{re}}_\mathcal{K}$. By construction, the domain of $\mathcal{I}^{\mathsf{re}}_\mathcal{K}$ is polynomial in the size of $\mathcal{K}$. Overall, this means that the rewriting procedure runs in polynomial time, and produces a polynomially bounded FO query.

Reduction to Classical DLs

After having considered the ontology-based query answering technique for rough DLs based on the combined approach in the last sections, we now take a brief look at a method for reducing this problem to QA in classical DLs that builds on proposals developed for rough DLs in the past.

It is known that rough DLs can be simulated in sufficiently expressive (classical) DLs (Schlobach, Klein, and Peelen 2007). Specifically, the upper and lower approximations $\underline{C}$ and $\overline{C}$ are equivalent to the concepts $\forall\rho.C$ and $\exists\rho.C$, respectively, where ρ is a designated transitive, reflexive, and symmetric role. Hence, one needs only to be able to express existential and value restrictions (as in the DL $\mathcal{ALC}$), and the three mentioned properties on roles. In other words, every rough-$\mathcal{EL}$ KB can be expressed by an $\mathcal{SI}^{\mathsf{Self}}$ KB.[2] Thus, any QA tool capable of dealing with this (very) expressive DL would also be able to handle rough $\mathcal{EL}$. Given the efforts to produce efficient QA methods for expressive DLs, one obvious question is whether such methods can be exploited directly to handle $\mathcal{ELH}^\rho_\bot$. The answer, unfortunately, is 'no'. The reason for this negative answer is that this logic does not fall into the class of Horn DLs, for which QA tools are efficient. In a nutshell, Horn DLs are those that do not allow the expression of non-deterministic choices (Ortiz, Rudolph, and Simkus 2011).

Recall the normal form for $\mathcal{ELH}^\rho_\bot$ TBoxes presented at the beginning of the last section. It is easy to see that, under the translation of Schlobach, Klein, and Peelen described at the beginning of this section, all the axioms in the first row are in fact Horn axioms. Unfortunately, this does not hold for the last axiom since it requires a value restriction on the left-hand side. This kind of constraint, which implicitly requires a non-deterministic choice (an element belongs to $\forall r.A$ if it either has no r-successors, or it has at least one r-successor, and all of them belong to A), cannot be handled efficiently by state-of-the-art QA tools.

On the other hand, the restriction of $\mathcal{ELH}^\rho_\bot$ where lower approximations cannot appear on the left-hand side of GCIs is, in fact, a Horn DL; more precisely, a sublanguage of Horn-$\mathcal{SI}^{\mathsf{Self}}$. Obviously, this restriction removes an important part of the expressive power of roughness, which may be fundamental for some practical applications. However, it is not hard to conceive cases where such lower approximations on the left-hand side are not really necessary. For instance, in our species classification and differentiation example, the TBox will fall within this sub-logic. Indeed, one may say that a property of a species is indiscernible from another, but a meaningful species description will never say that an element that is indiscernible from all in a species must satisfy some specific properties.

There are approaches for conjunctive query answering that extend $\mathcal{EL}$ directly towards the expressivity needed for rough $\mathcal{EL}$. For instance, in (Stefanoni and Motik 2015) the authors investigate an extension of $\mathcal{EL}$ that allows for reflexive and transitive roles, but not for symmetric ones, which in general damage the tractability of $\mathcal{EL}$. Their techniques were implemented in the system RDFox (Motik et al. 2014). As mentioned, this DL covers two of the three properties of an equivalence relation. Symmetry for roles is missing in their approach, since symmetric roles behave to some ex-

[2] $\mathcal{SI}^{\mathsf{Self}}$ extends $\mathcal{ALC}$ with transitive and inverse roles, and reflexivity statements. For more details, see (Baader et al. 2007).

tent similarly to inverse roles which are notorious for raising the computational complexity of reasoning in many logics. Even transitive roles alone are known to be a handicap to the performance of query answering systems for $\mathcal{EL}$ including them. However, for $\mathcal{EL}$ with transitive roles practical reasoning procedures based on the combined approach have been devised in (Lutz et al. 2013) and implemented in the Combo system.

Conclusions

We have presented a combined approach for answering conjunctive queries in the rough DL $\mathcal{ELH}^{\rho}_{\bot}$. This approach first extends the input ABox to include also the knowledge encoded in the TBox by materialization, and then rewrites the query to guarantee that no answers are unexpectedly introduced in the first step. This allows us to effectively answer conjunctive queries in this rough DL using standard database technologies.

Interestingly, we have shown that dealing with this rough extension of $\mathcal{ELH}_{\bot}$ does not incur in any increase of complexity w.r.t. its classical counterpart; the rewriting remains polynomial in the size of the input.

Being able to model and reason with rough concepts is fundamental for applications in the life sciences, as they allow the introduction of notions that cannot be precisely defined through use of approximating lower and upper bounds. In addition, they allow to introduce examples of elements that cannot be distinguished by these approximations. Such approaches have recently been investigated for a more fine-grained setting, where vagueness can be captured by a similarity measure and a proto-typical instance, yielding a vague concept that can be dynamically relaxed or strengthened depending on a similarity threshold (Baader, Brewka, and Fernández Gil 2015)—albeit only for unfoldable TBoxes. In our setting the query language itself allows to relax answers by admitting the indiscernibility relation and the approximation constructors in the query language. Here the degree of relaxation then depends on the presence of the indiscernibility relation in the data. A somewhat orthogonal approach has been investigated in (Ecke, Peñaloza, and Turhan 2015), where the query language admits relaxation (of instance queries) by the use of a concept similarity measure and a threshold. While the similarity-based approaches admit more flexibility, they crucially depend on the presence of an appropriate similarity measure supplied by the user. In case of approaches using rough DLs, the indiscernibility relation can, in principle, be automatically derived from the data (d'Amato et al. 2013; Beek, Schlobach, and van Harmelen 2016).

We highlight that there exist database systems providing native support for rough sets (Hu, Lin, and Han 2004; Beer and Bühler 2015). As an alternative approach, one could think of using them as a target language for rewriting the queries. While this would solve some of the technical issues regarding indiscernible elements in the query rewriting step, these systems are not as widely adopted and optimized as industrial database systems. Hence we believe that our approach has a higher potential for practical impact.

We plan to implement the rewriting technique and to test its performance empirically. We will also extend our methods to weaker notions of roughness, by removing restrictions in the indiscernibility relation; e.g. transitivity.

Acknowledgments

This work is partly supported by the German Research Foundation (DFG) within the Cluster of Excellence 'Center for Advancing Electronics Dresden' (cfaed) in CRC 912 (HAEC) and within the DFG project 'Reasoning and Query Answering Using Concept Similarity Measures and Graded Membership Functions' BA 1122/20-1.

References

Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F., eds. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2nd edition.

Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the $\mathcal{EL}$ envelope. In *Proc. of 18th Int. Joint Conference on Artificial Intelligence (IJCAI 2005)*.

Baader, F.; Brewka, G.; and Fernández Gil, O. 2015. Adding threshold concepts to the description logic $\mathcal{EL}$. In *Proceedings of the 10th International Symposium on Frontiers of Combining Systems (FroCoS'15)*, volume 9322 of *LNAI*, 33–48. Springer.

Beek, W.; Schlobach, S.; and van Harmelen, F. 2016. A contextualised semantics for owl: sameAs. In *Proceedings of ESWC*, volume 9678 of *LNCS*, 405–419. Springer.

Beer, F., and Bühler, U. 2015. An in-database rough set toolkit. In *Proc. of the LWA 2015 Workshops*, volume 1458 of *CEUR Workshop Notes*, 146–157.

Bobillo, F.; Cerami, M.; Esteva, F.; García-Cerdaña, À.; Peñaloza, R.; and Straccia, U. 2015. Fuzzy description logic. In Cintula, P.; Fermüller, C. G.; and Noguera, C., eds., *Handbook of Mathematical Fuzzy Logic Volume 3*, volume 58 of *Studies in Logic*. College Publications.

Borgwardt, S.; Cerami, M.; and Peñaloza, R. 2017. The complexity of fuzzy $\mathcal{EL}$ under the Łukasiewicz t-norm. *Int. J. Approx. Reasoning* 91:179–201.

Borgwardt, S.; Distel, F.; and Peñaloza, R. 2015. The limits of decidability in fuzzy description logics with general concept inclusions. *Artificial Intelligence* 218:23–55.

d'Amato, C.; Fanizzi, N.; Esposito, F.; and Lukasiewicz, T. 2013. Representing uncertain concepts in rough description logics via contextual indiscernibility relations. In *Int. Workshop on Uncertainty Reasoning for the Semantic Web*, volume 7123 of *LNCS*, 300–314. Springer.

Ecke, A.; Peñaloza, R.; and Turhan, A.-Y. 2015. Similarity-based relaxed instance queries. *Journal of Applied Logic* 13(4, Part 1):480–508. Special Issue for the Workshop on Weighted Logics for AI 2013.

Hu, X.; Lin, T. Y.; and Han, J. 2004. A new rough sets model based on database systems. *Fundam. Inform.* 59(2-3):135–152.

Jiang, Y.; Wang, J.; Tang, S.; and Xiao, B. 2009. Reasoning with rough description logics: An approximate concepts approach. *Inf. Sci.* 179(5):600–612.

Keet, C. M. 2010. Ontology engineering with rough concepts and instances. In *Proc. of 17th International Conference on Knowledge Engineering and Management by the Masses EKAW 2010*, volume 6317 of *LNCS*, 503–513. Springer.

Keet, C. M. 2011. Rough subsumption reasoning with rOWL. In *Proc. of the 2011 Ann. Conf. of the South African Inst. of Computer Scientists and Information Technologists, SAICSIT 2011*, 133–140. ACM.

Klein, M. C.; Mika, P.; and Schlobach, S. 2007. Rough description logics for modeling uncertainty in instance unification. In *Proc. of 3rd ISWC Workshop on Uncertainty Reasoning for the Semantic Web*, volume 327 of *CEUR Workshop Notes*.

Liau, C.-J. 1996. On rough terminological logics. In *Proc. of the 4th Int. Workshop on Rough Sets, Fuzzy Sets and machine Discovery (RSFD'96)*, 47–54.

Lin, T. Y., and Cercone, N. 2012. *Rough sets and data mining: Analysis of imprecise data*. Springer Science & Business Media.

Lukasiewicz, T., and Straccia, U. 2008. Managing uncertainty and vagueness in description logics for the semantic web. *J. Web Sem.* 6(4):291–308.

Lutz, C., and Wolter, F. 2010. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *Journal of Symbolic Computation* 45(2):194–228.

Lutz, C.; Seylan, I.; Toman, D.; and Wolter, F. 2013. The combined approach to OBDA: taming role hierarchies using filters. In *Proc. of the 12th Int. Semantic Web Conference ISWC 2013*, volume 8218 of *LNCS*, 314–330. Springer.

Lutz, C.; Toman, D.; and Wolter, F. 2009. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *Proc. of 20th Int. Joint Conference on Artificial Intelligence (IJCAI 2009)*, 2070–2075.

Motik, B.; Nenov, Y.; Piro, R.; Horrocks, I.; and Olteanu, D. 2014. Parallel materialisation of datalog programs in centralised, main-memory RDF systems. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence.*, 129–137. AAAI Press.

Ortiz, M.; Rudolph, S.; and Simkus, M. 2011. Query answering in the Horn fragments of the description logics $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$. In *Proc. of the 22nd Int. Joint Conference on Artificial Intelligence (IJCAI 2011)*, 1039–1044. IJCAI/AAAI.

Pawlak, Z. 1982. Rough sets. *International Journal of Parallel Programming* 11(5):341–356.

Pawlak, Z. 1998. Reasoning about data - A rough set perspective. In *Proc. of First Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'98)*, volume 1424 of *LNCS*, 25–34. Springer.

Peñaloza, R., and Zou, T. 2013. Roughening the $\mathcal{EL}$ envelope. In *Proc. of Int. Symposium on Frontiers of Combining Systems (FroCoS 2013)*, volume 8152 of *LNCS*, 71–86. Springer.

Peñaloza, R.; Thost, V.; and Turhan, A.-Y. 2018. Query answering for rough $\mathcal{EL}$ ontologies (extended technical report). *CoRR* abs/1808.01877.

Schlobach, S.; Klein, M. C.; and Peelen, L. 2007. Description logics with approximate definitions - precise modeling of vague concepts. In *Proc. of 19th Int. Joint Conference on Artificial Intelligence (IJCAI 2007)*, 557–562.

Stefanoni, G., and Motik, B. 2015. Answering conjunctive queries over $\mathcal{EL}$ knowledge bases with transitive and reflexive roles. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 1611–1617. AAAI Press.