Counter Model Transformation for Explaining Non-Subsumption in $\mathcal{EL}$

Christian Alrabbaa, Willi Hieke, and Anni-Yasmin Turhan

Institute for Theoretical Computer Science
TU Dresden, Dresden, Germany
`firstname.lastname@tu-dresden.de`

Abstract. When subsumption relationships unexpectedly fail to be detected by a description logic reasoner, the cause for this "non-entailment" need not be evident. In this case, succinct automatically generated explanations would be helpful. Reasoners for the description logic $\mathcal{EL}$ compute the canonical model of a TBox in order to perform subsumption tests. We devise parts of such models as relevant parts for explanation and propose an approach based on graph transductions to extract such relevant parts from canonical models.

Keywords: Explainable AI · Description Logic · Model Transformation.

1 Introduction

Description logics (DLs) are a family of decidable fragments of first-order logic (FOL) that can speak about graph structures. These logics are commonly used to formalize knowledge about concepts about the real world. Concepts in description logics are predicates built from unary and binary FOL predicates. Knowledge about concepts is then expressed in an ontology, which is a finite set of concept relating axioms. In the case of DLs, ontologies are referred to as TBoxes. DLs are well investigated w.r.t. to the complexity of their reasoning problems. The practical relevance of DLs becomes apparent as they provide the formal basis for W3C standardized ontologies comprised in the OWL 2 standard. The corresponding reasoning procedures are implemented in highly optimized reasoner systems [8]. DL ontologies from practical applications easily contain more than 10.000s of axioms. Therefore, it sometimes is not obvious why an expected consequence does not hold. When users with little expertise in logic face such a situation, automated explanation services are needed. A common reasoning problem for DL TBoxes is to decide whether a subsumption relationship holds between two given concepts, i.e., to decide whether the first concept is a specialization of the second concept w.r.t. a TBox. Decision procedures for subsumption have been implemented in a range of DL reasoners [8]. In this paper, we consider the setting where the non-consequence in question is a missing subsumption relationship between two concepts w.r.t. an $\mathcal{EL}$ TBox. The DL $\mathcal{EL}$ is computationally well-behaved as deciding subsumption is tractable [5]. Also, $\mathcal{EL}$

enjoys the canonical model property, which guarantees the existence of a particular standardized model with useful properties. Reasoning in $\mathcal{EL}$ then amounts to computing the canonical model. In fact, many $\mathcal{EL}$ reasoners implement the computation of the canonical model [3,10,11]. Explaining negative answers to a subsumption question can be addressed by supplying counter examples. Such counter examples could then either be displayed to the user or serve as a starting point for generating more user-friendly explanations. A counter example for non-subsumption is the canonical model of the TBox itself. As this model contains the whole signature of the TBox, it can easily be too large for explanation purposes. The crucial step is to identify relevant parts of the counter model that are useful for explaining the non-entailment. We propose in this paper four kinds of relevant substructures of canonical models that can serve as explanations and propose an approach to extract these parts from canonical models. As formalisms to specify the extraction of relevant parts from a canonical model, we make us of transductions. In general, transductions specify mappings over relational structures and have been used to transform models of TBoxes in [9].

For explaining positive answers to a subsumption test, early approaches use justifications as explanations, which are subset-minimal subsets of the TBox that are "responsible" for the subsumption [4,13]. These methods produce syntactic explanations using axioms that appear in the TBox. Explaining positive subsumption results can also be done by providing proof, i.e., a derivation of the subsumption by a calculus. Such methods were recently investigated for the case of the DL $\mathcal{EL}$ in [1]. Our approach to explaining non-subsumption is more fine-grained than classical justification-based methods in the sense that it can address consequences of the TBox individually, since it uses the semantics of the DL KB. Furthermore, unlike proof-based techniques, the outcome of our method is of declarative nature and thus suits the declarative nature of DLs.

2 Preliminaries

Description logics are a family of decidable knowledge representation formalisms that can model structures over unary and binary predicates. Unary predicates are called *concepts* and binary predicates are called *roles*. Concepts for the description logic $\mathcal{EL}$ are built inductively from a set of concept names $\mathsf{N_A}$ and a set of role names $\mathsf{N_R}$. Let $A \in \mathsf{N_A}$ and $r \in \mathsf{N_R}$, then (complex) $\mathcal{EL}$-concepts are built by the syntactic rule:

$$C ::= A \mid C \sqcap C \mid \exists r.C \mid \top.$$

We assume that the sets of concept names and role names are disjoint. We denote concepts by upper case and roles with lower case letters. For concept names, we usually write A and B, and by C and D we indicate possibly complex concepts. A signature Σ is a union of two finite sets $\Sigma_{\mathsf{C}} \subset \mathsf{N_A}$ and $\Sigma_{\mathsf{R}} \subset \mathsf{N_R}$.

An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ over a signature Σ consists of a non-empty set $\Delta^{\mathcal{I}}$ called the *interpretation domain* and an *interpretation function* $\cdot^{\mathcal{I}}$ that maps every concept name in Σ_C to a subset of $\Delta^\mathcal{I}$ and every role name in Σ_R to a subset of $\Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The mapping $\cdot^\mathcal{I}$ extends to concepts:

- $(C \sqcap D)^\mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I}$,
- $(\exists r.C)^\mathcal{I} = \{d \in \Delta^\mathcal{I} \mid \text{there is an } e \in C^\mathcal{I} \textit{ s.t. } (d,e) \in r^\mathcal{I}\}$,
- and $\top^\mathcal{I} = \Delta^\mathcal{I}$.

An $\mathcal{EL}$ *TBox* $\mathcal{T}$ is a finite set of *concept inclusions (CIs)*, which are formulae of the form $C \sqsubseteq D$, where C and D are $\mathcal{EL}$ concepts. We abbreviate $C \sqsubseteq D$ and $D \sqsubseteq C$ by $C \equiv D$. An interpretation $\mathcal{I}$ satisfies a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$. An interpretation $\mathcal{I}$ is called a *model* of a TBox $\mathcal{T}$ if $\mathcal{I}$ satisfies all the CIs in $\mathcal{T}$. For a TBox, an interpretation or a concept X, we denote its *signature* by $\mathsf{sig}(X)$. To denote the *concept signature of* X, we write use $\mathsf{sig}_\mathsf{C}(X) = \mathsf{sig}(X) \cap \mathsf{N_C}$, and $\mathsf{sig}_\mathsf{R}(X) = \mathsf{sig}(X) \cap \mathsf{N_R}$ for the *role signature of* X. A prominent reasoning problem for DLs is to decide subsumption. Given two concepts C and D and a TBox $\mathcal{T}$, *subsumption* (denoted $C \sqsubseteq_\mathcal{T} D$) decides whether for each model $\mathcal{I}$ of $\mathcal{T}$, $C^\mathcal{I} \subseteq D^\mathcal{I}$ holds. Since $\mathcal{EL}$ cannot express negation, satisfiability is trivial. The main method for deciding subsumption is to compute the canonical model of an $\mathcal{EL}$ TBox [5]. The method to compute the canonical model first normalizes the TBox. An $\mathcal{EL}$ TBox is in *normal form* if and only if it only contains CIs of the forms:

$$A \sqsubseteq B, \qquad A_1 \sqcap A_2 \sqsubseteq B, \qquad \exists r.A \sqsubseteq B, \qquad \textit{or} \qquad A \sqsubseteq \exists r.B,$$

where A, A_1, A_2, and B are concept names or $\top$, and r is a role name. Every $\mathcal{EL}$ TBox $\mathcal{T}$ can be transformed into a TBox $\mathcal{T}'$ in normal form such that the size of $\mathcal{T}'$ is linear in the size of $\mathcal{T}$ and every model of $\mathcal{T}'$ is a model of $\mathcal{T}$ [2].

Definition 1 (Canonical Model [2]). *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox. The* canonical model $\mathcal{I}_\mathcal{T}$ *of $\mathcal{T}$ is defined as follows:*

$$\begin{aligned}
\Delta^{\mathcal{I}_\mathcal{T}} &:= \{A \mid A \in \mathsf{N_C} \cap \mathsf{sig}(\mathcal{T})\} \cup \{\top\},\\
A^{\mathcal{I}_\mathcal{T}} &:= \{B \in \Delta^{\mathcal{I}_\mathcal{T}} \mid B \sqsubseteq_\mathcal{T} A\} \textit{ for all } A \in \Sigma_\mathsf{C},\\
r^{\mathcal{I}_\mathcal{T}} &:= \{(A,B) \in \Delta^{\mathcal{I}_\mathcal{T}} \times \Delta^{\mathcal{I}_\mathcal{T}} \mid A \sqsubseteq_\mathcal{T} \exists r.B\} \textit{ for all } r \in \Sigma_\mathsf{R}.
\end{aligned}$$

For any normalized $\mathcal{EL}$ TBox $\mathcal{T}$, its canonical model $\mathcal{I}_\mathcal{T}$, and for two named concepts A and B, we have that $A \sqsubseteq_\mathcal{T} B$ if and only if $\mathcal{I}_\mathcal{T} \models A \sqsubseteq B$ [2]. Subsumption of arbitrary concepts C and D can be tested, if the CIs: $A \sqsubseteq C$ and $B \sqsubseteq D$ are added to $\mathcal{T}$. In a canonical model, each concept name from the (normalized) TBox is represented by one element in the domain. An element $a \in A^{\mathcal{I}_\mathcal{T}}$ is *the representative of A in $\mathcal{I}_\mathcal{T}$* if for all other $x \in A^{\mathcal{I}_\mathcal{T}}$ there is a concept name N, s.t. $x \in N^{\mathcal{I}_\mathcal{T}}$ and $a \notin N^{\mathcal{I}_\mathcal{T}}$.

Model transformations is a binary (usually functional) relations on the class of finite DL interpretations. We mostly use monadic second-order (MSO) transductions as formalism to describe model transformations. Such transductions are defined in [7] and tailored to description logic interpretations in [9]. Intuitively, a transduction specifies a model transformation by a tuple of MSO formulae, called *definition scheme*, that describes how to construct an output interpretation in terms of the input interpretation. We will use transductions to manipulate models of $\mathcal{EL}$ TBoxes and hence only deal with binary relational signatures. We denote these signatures by Σ and write Σ_C for the set of its unary predicates and Σ_R for the set of its binary predicates. By $\mathsf{MSO}(\Sigma, \mathcal{W})$ we denote the set of MSO formulae with free first-order variables in $\mathcal{W}$. These variables are called *parameters*.

Definition 2 (Monadic Second-Order Definition Scheme). *Let Σ be a binary signature, and let $\mathcal{W}$ be a finite set of* parameters. *A* monadic second-order definition scheme *is a tuple*

$$\mathsf{D} = \langle \chi, \delta, (\theta_N)_{N \in \Sigma_\mathsf{C}}, (\eta_r)_{r \in \Sigma_\mathsf{R}} \rangle, \ \textit{where}$$

- $\chi \in MSO(\Sigma, \mathcal{W})$ *is called the* precondition*,*
- $\delta \in MSO(\Sigma, \mathcal{W} \cup \{x\})$ *is called the* domain formula*,*
- $\theta_N \in MSO(\Sigma, \mathcal{W} \cup \{x\})$ *for all $N \in \Sigma_\mathsf{C}$ are the* concept formulae*,*
- $\eta_r \in MSO(\Sigma, \mathcal{W} \cup \{x, y\})$ *for all $r \in \Sigma_\mathsf{R}$ are the* role formulae*.*

The different formulae serve different purposes. The precondition χ needs to be satisfied by the input interpretation. This is essentially a test whether the transformation specified by the scheme is applicable to the input. The *domain formula* δ defines the interpretation domain of the output interpretation. More precisely, it selects those elements from the input interpretation that satisfy δ. For these domain elements, the *concept formulae* θ and *role formulae* η define the extensions of the named concept and roles for each symbol from the signature. MSO definition schemes are employed to generate transductions on interpretations. For our purpose of extracting relevant parts of models, we can restrict ourselves to the case of non-copying transductions, i.e. to transductions that do not increase the size of the interpretation.

Definition 3 (Transduction Induced by a Definition Scheme). *Let $\mathcal{I}$ be an interpretation over a binary signature Σ, let $\mathcal{W}$ be a set of parameters, and let λ be a $\mathcal{W}$-assignment in $\mathcal{I}$, i.e., $\lambda : \mathcal{W} \to \Delta^\mathcal{I}$. A definition scheme D defines the interpretation $\mathcal{I}'$ from $(\mathcal{I}, \lambda)$ if*

- $(\mathcal{I}, \lambda) \models \chi(\mathcal{W})$,[1]
- $\Delta^{\mathcal{I}'} := \{a \in \Delta^\mathcal{I} \mid (\mathcal{I}, \lambda) \models \delta(\mathcal{W}, a)\}$,
- $A^{\mathcal{I}'} := \{a \in \Delta^{\mathcal{I}'} \mid (\mathcal{I}, \lambda) \models \theta_N(\mathcal{W}, a)\}$ *for all* $N \in \Sigma_\mathsf{C}$,
- $r^{\mathcal{I}'} := \{(a, b) \in (\Delta^{\mathcal{I}'})^2 \mid (\mathcal{I}, \lambda) \models \eta_r(\mathcal{W}, a, b)\}$ *for all* $r \in \Sigma_\mathsf{R}$,

with $(\mathcal{I}, \lambda) \models \delta(\mathcal{W}, a)$ meaning $(\mathcal{I}, \lambda') \models \delta(\mathcal{W}, x)$, where λ' is the assignment extending λ such that $\lambda' : x \mapsto a$ (and accordingly for θ and η). We denote $\widehat{\mathsf{D}}(\mathcal{I}, \lambda) = \mathcal{I}'$. The transduction τ *induced by* D *is defined as*

$$\tau := \{(\mathcal{I}, \widehat{\mathsf{D}}(\mathcal{I}, \lambda)) \mid \lambda \textit{ is a } \mathcal{W}\textit{-assignment in } \mathcal{I} \textit{ with } (\mathcal{I}, \lambda) \models \chi\},$$

and $\tau(\mathcal{I})$ denotes $\{\widehat{\mathsf{D}}(\mathcal{I}, \lambda) \mid (\mathcal{I}, \lambda) \models \chi$ for some $\lambda\}$. For functional transductions we write $\tau(\mathcal{I}) = \mathcal{I}'$.

[1] We indicate parameters variables from the set $\mathcal{W} = \{z_1, \ldots, z_n\}$ by writting $\delta(\mathcal{W}, x)$ instead of $\delta(z_1, \ldots, z_n, x)$.

3 Defining Relevant Parts of Counter Models

Recall that we want to explain a non-subsumption by showing relevant parts of a counter model. We start from the basic definition of a counter model. We refer to asking for the validity of a subsumption relation ϕ in a given TBox as *subsumption query.* Let $\mathcal{T}$ be a TBox and $\phi := A \sqsubseteq B$ a subsumption query that uses w.l.o.g. named concepts, s.t. $\mathcal{T} \not\models \phi$. An interpretation $\mathcal{I}$ is called a *counter model for ϕ w.r.t. $\mathcal{T}$* iff $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \not\models \phi$. Presenting an entire model of the full knowledge base to the user is not ideal because it contains a lot of irrelevant parts. In this section, we provide four different definitions of what relevant parts of a counter model are. The goal is to reduce the amount of information in the counter model, i.e., reducing the amount of domain elements, concept labels, and role labels in order to provide a concise explanation of the non-subsumption. In case of the DL $\mathcal{EL}$, the canonical model $\mathcal{I}_\mathcal{T}$ is the standard counter model and we describe our methods for this kind of model. We exemplify our approach by a running example.

Example 4. Clinical differentiation between Parkinson's disease (PD) and progressive supranuclear palsy (PSP) can be challenging due to overlapping clinical features [12]. We model the characteristics of patients with the diseases by the following TBox.

$$
\begin{aligned}
\mathcal{T}_{ex} := \{ \quad \mathsf{PD} &\sqsubseteq \mathsf{NeuroDisease} \sqcap \exists \mathsf{accumulates}.\mathsf{AlphaProtein} \sqcap \\
&\quad \exists \mathsf{lossOf}.\mathsf{Mobility} \sqcap \exists \mathsf{has}.\mathsf{Tremor}, \\
\mathsf{PSP} &\sqsubseteq \mathsf{NeuroDisease} \sqcap \exists \mathsf{accumulates}.\mathsf{TauProtein} \sqcap \\
&\quad \exists \mathsf{lossOf}.\mathsf{Mobility} \sqcap \exists \mathsf{impairs}.\mathsf{Speech}, \\
\mathsf{PDPatient} &\sqsubseteq \exists \mathsf{diagnosedWith}.\mathsf{PD}, \quad \mathsf{TauProtein} \sqsubseteq \mathsf{Protein} \sqcap \exists \mathsf{builds}.\mathsf{Tubuli}, \\
\mathsf{PSPPatient} &\sqsubseteq \exists \mathsf{diagnosedWith}.\mathsf{PSP}, \quad \mathsf{AlphaProtein} \sqsubseteq \mathsf{Protein} \quad \}
\end{aligned}
$$

Our example subsumption query is $\phi_{ex} := \mathsf{PD} \sqsubseteq_{\mathcal{T}_{ex}} \mathsf{PSP}$ which is not entailed by $\mathcal{T}_{ex}$ and for which we want to supply relevant parts from the canonical model of $\mathcal{T}_{ex}$ as this is our standard counter model. Figure 1 depicts the canonical model $\mathcal{I}_{\mathcal{T}_{ex}}$ of $\mathcal{T}_{ex}$ using the obvious abbreviations for the names. $\mathcal{I}_{\mathcal{T}_{ex}}$ contains element a as the representative for the concept PD and b for the concept PSP.

We want to identify relevant substructures of a counter model by requiring that these substructures to be models of sets of implications that follow from the TBox $\mathcal{T}$, and in that sense preserve parts of the model. Therefore, we define sets of implications w.r.t. a given TBox $\mathcal{T}$. With $\mathcal{EL}(\mathsf{sig}(\mathcal{T}))$ denoting $\mathcal{EL}$ concepts written in the signature of $\mathcal{T}$, we define

$$\mathsf{Sub}_\mathcal{T}(C) := \{H \mid C \sqsubseteq_\mathcal{T} H, H \in \mathcal{EL}(\mathsf{sig}(\mathcal{T}))\}.$$

By $H[{}^{N}\!/\!{}_{M}, {}^{R}\!/\!{}_{S}]$ we denote the exhaustive syntactic substitution of first, every occurrence of N by M and second, every occurrence of R by S.

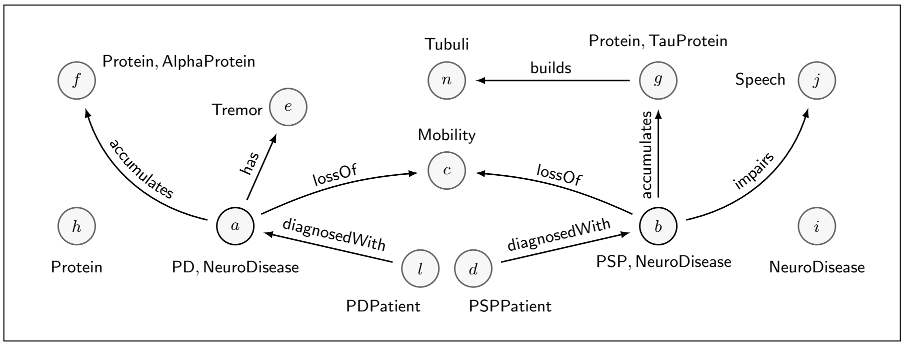


Fig. 1: Canonical model $\mathcal{I}_{\mathcal{T}_{ex}}$ of TBox $\mathcal{T}_{ex}$.

Definition 5 (Relevant Implication Sets). *Let $\mathcal{T}$ be a TBox and $A, B \in \mathsf{sig}(\mathcal{T})$ be concept names. The* relevant implication sets of $\mathcal{T}$ *w.r.t. A and B are:*

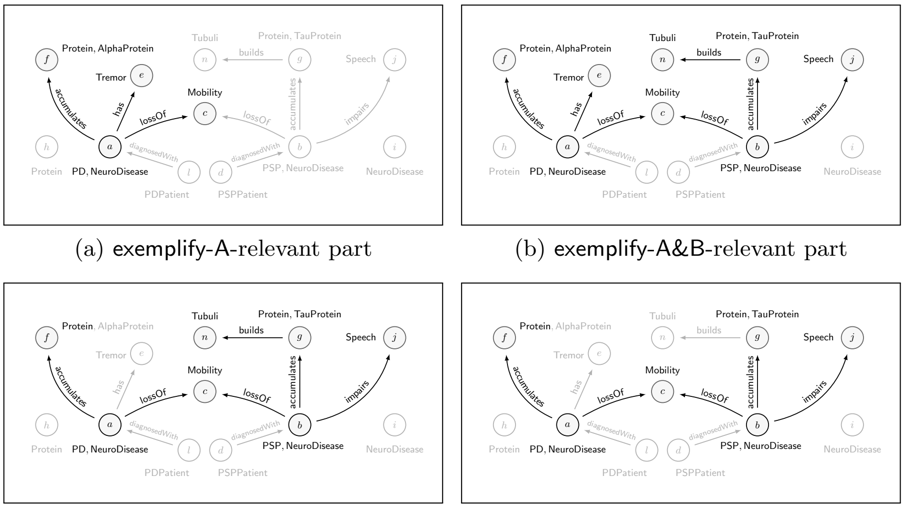
$$\begin{aligned}
\mathrm{S}_{\mathcal{T}}(A) &:= \{A \sqsubseteq H \mid H \in \mathit{Sub}_{\mathcal{T}}(A)\},\\
\mathrm{C}_{\mathcal{T}}(A,B) &:= \{G \sqsubseteq H \mid H \in \mathit{Sub}_{\mathcal{T}}(A) \cap \mathit{Sub}_{\mathcal{T}}(B), G \in \{A,B\}\},\\
\bar{\mathrm{S}}_{\mathcal{T}}(A,B) &:= \{B \sqsubseteq H \mid H \in \mathit{Sub}_{\mathcal{T}}(B), H[{}^{N}/_{\top}, {}^{\exists r.\top}/_{\top}] \in \mathit{Sub}_{\mathcal{T}}(A)\\
&\qquad \textit{for all } N \in \mathsf{sig}(H) \cap \Sigma_{\mathsf{C}} \textit{ and all } r \in \mathsf{sig}(H) \cap \Sigma_{\mathsf{R}}\}.
\end{aligned}$$

Intuitively, $\mathrm{S}_{\mathcal{T}}(A)$ consists of all the CIs that preserve the information on the instances of A w.r.t. $\mathcal{T}$. The set $\mathrm{C}_{\mathcal{T}}(A,B)$ contains CIs that preserve for A and for B the information on the commonalities of A and B w.r.t. $\mathcal{T}$. The CIs in $\bar{\mathrm{S}}_{\mathcal{T}}(A,B)$ preserve for B some commonalities of A and B that follow from $\mathcal{T}$. These commonalities are restricted to those subsumers of A that remain subsumers, if all concept names are removed from them and the role-depth of each nested existential restriction is reduced by 1. The idea will become evident with the continuation of the running example.

Definition 6 (Relevant Parts of Counter Models). *Let $\mathcal{T}$ be a TBox, $\phi := A \sqsubseteq_{\mathcal{T}} B$ a subsumption query, $\mathcal{I}$ a counter model of ϕ w.r.t. $\mathcal{T}$. Let $a, b \in \Delta^{\mathcal{I}}$ s.t. $a \in A^{\mathcal{I}} \setminus B^{\mathcal{I}}$, and $b \in B^{\mathcal{I}}$ if $B^{\mathcal{I}} \neq \emptyset$. An interpretation $\mathcal{I}'$ is called* {*exemplify-A*, *exemplify-A&B*, *diff*, *flat-diff*}-relevant part of $\mathcal{I}$ w.r.t. ϕ and $\mathcal{T}$ *iff $\mathcal{I}'$ is one of the smallest substructures of $\mathcal{I}$, s.t. $\mathcal{I}' \not\models \phi$, and either*

- $a \in A^{\mathcal{I}'}$ *and* $\mathcal{I}' \models \mathrm{S}_{\mathcal{T}}(A)$; (exemplify-A)
- $a \in A^{\mathcal{I}'}, b \in B^{\mathcal{I}'}$ *and* $\mathcal{I}' \models \mathrm{S}_{\mathcal{T}}(A) \cup \mathrm{S}_{\mathcal{T}}(B)$; (exemplify-A&B)
- $a \in A^{\mathcal{I}'}, b \in B^{\mathcal{I}'}$ *and* $\mathcal{I}' \models \mathrm{C}_{\mathcal{T}}(A,B) \cup \mathrm{S}_{\mathcal{T}}(B)$; (diff)
- $a \in A^{\mathcal{I}'}, b \in B^{\mathcal{I}'}$ *and* $\mathcal{I}' \models \mathrm{C}_{\mathcal{T}}(A,B) \cup \bar{\mathrm{S}}_{\mathcal{T}}(A,B,)$. (flat-diff)

By smallest substructure $\mathcal{I}'$ of $\mathcal{I}$ w.r.t. ϕ and $\mathrm{S}_{\mathcal{T}}(A)$, we mean that any strict substructure $\mathcal{I}''$ of $\mathcal{I}'$ is not a model of ϕ and $\mathrm{S}_{\mathcal{T}}(A)$ anymore — and likewise for the other relevance notions.



(a) exemplify-A-relevant part

(b) exemplify-A&B-relevant part

(c) diff-relevant part

(d) flat-diff-relevant part

Fig. 2: The different kinds of relevant parts of the canonical model $\mathcal{I}_{ex}$ of $\mathcal{T}_{ex}$.

We explain the intuition and the purpose of the four relevance parts in reference to Figure 2, where the corresponding parts of the canonical model $\mathcal{I}_{\mathcal{T}_{ex}}$ from our running example are depicted. The exemplify-A-relevant part, depicted in Figure 2a, highlights the information on A w.r.t. $\mathcal{T}$ and hence can be used to display a "full" example of an instance of A from the subsumption query that is not subsumed by B. The exemplify-A&B-relevant part, displayed in Figure 2b, follows the same idea but for both concepts A and B. Therefore, the "full" information from both query concepts can be shown to the user. These two kinds of relevant parts give a full descriptions of the involved concepts. Explaining non-entailment can be considered as a kind of abduction problem, which, in our case, is to infer what A lacks compared to B. This consideration motivates the remaining two relevance parts we suggest. The diff-relevant part shows the difference of A and B by preserving the information on the commonalities of both concepts at a and gives full information on B at b. It thereby highlights which parts of B are not entailed for A. The diff-relevant part is displayed in Figure 2c. The flat-diff-relevant part illustrates a flattened form of difference as it preserves only those parts from B up to the smallest depth where a difference to A occurs. The flat-diff-relevant part in Figure 2d prunes the relational structure of b in comparison to the diff-relevant part. In particular, diff shows that PSP accumulates tau proteins and also that they build tubuli, whereas flat-diff merely shows that PSP accumulates tau proteins. That is already is a difference to PD, which accumulates alpha protein. The fact that tau protein build tubuli is hidden because it already suffices to know that PSP requires the accumulations of tau proteins to explain the difference between PD and PSP.

4 Extracting Relevant Parts of Counter Models

Given a canonical model of a TBox and a subsumption query, one can obtain the four kinds of counter model parts from Definition 5 by model transformation. As input for the transformation we consider canonical models of normalized TBoxes, which are finite for general $\mathcal{EL}$ TBoxes and computable in polynomial time [2]. For the remainder of the paper we assume, that the subsumption query is $\phi := A \sqsubseteq_{\mathcal{T}} B$ w.r.t. a normalized TBox $\mathcal{T}$. To devise the transductions, we use the predicate $reach(x, y)$ for reachability of two elements in an interpretation $\mathcal{I}$ [7, Section 5.2.2]. An edge between two elements of an interpretation $\mathcal{I}$ is defined by $succ(x, y) := \bigvee_{r \in \Sigma_{\mathsf{R}}} r(x, y)$. We introduce the definition scheme $\mathsf{D}_{\text{exemplify-A}}$, that induces transduction $\tau_{\text{exemplify-A}}$. The idea is that the transduction $\tau_{\text{exemplify-A}}$ extracts the exemplify-A-relevant part from the canonical model $\mathcal{I}_{\mathcal{T}}$ of the $\mathcal{EL}$ TBox.

Definition 7 (Transduction $\tau_{\text{exemplify-A}}$). *Let Σ be a signature s.t. $A, B \in \Sigma_{\mathsf{C}} \subset \Sigma$, and let $\mathcal{W} = \{u\}$. The definition scheme $\mathsf{D}_{\text{exemplify-A}}$ inducing the transduction $\tau_{\text{exemplify-A}}$ consists of the formulae:*

$$\chi(\mathcal{W}) := A(u) \wedge \neg B(u), \qquad \theta_N(\mathcal{W}, x) := N(x),$$
$$\delta(\mathcal{W}, x) := reach(u, x), \qquad \eta_r(\mathcal{W}, x, y), := r(x, y).$$

We assume that the parameter u is assigned by $\lambda(u)$ to the representative for A as the expected subsumee from the subsumption query in $\mathcal{I}_{\mathcal{T}}$. We use this assumption also for the subsequent definition schemes. The preconditions indicate which parameter is mapped to which concept representative. The transduction simply picks the representative of A as the witness for the non-subsumption $A \sqsubseteq_{\mathcal{T}} B$ from $\mathcal{I}_{\mathcal{T}}$ and collects all the reachable successors of it to induce a substructure of $\mathcal{I}_{\mathcal{T}}$ w.r.t. $\mathsf{sig}(\mathcal{I}_{\mathcal{T}})$. Since the representative is unique for canonical models, we treat this transduction as functional.

Lemma 8. *Let Γ be a set containing only $\mathcal{EL}$ CIs of the form $A \sqsubseteq H$, where A is a concept name and H an arbitrary $\mathcal{EL}$ concepts. Let $\mathcal{I}$ be a model of Γ with $a \in A^{\mathcal{I}}$. Then, the induced substructure $\mathcal{I}'$ of $\mathcal{I}$ w.r.t. $\mathsf{sig}(\mathcal{I})$ that contains only the element a and all elements reachable from a, is a model of Γ.*

Proof. We show $(\mathcal{I}', a) \models H$ for all $\mathcal{EL}$ concepts H by induction on the structure of H. For the induction base, we assume $H = A'$ for $A' \in \mathsf{N_C}$. The claim holds since we have that $a \in \Delta^{\mathcal{I}'}$ and that $a \in C^{\mathcal{I}'}$ for every concept name C if and only if $a \in C^{\mathcal{I}}$ because $\mathcal{I}'$ is an induced substructure of $\mathcal{I}$, meaning that b get the very same concept labels in $\mathcal{I}'$ as it has in $\mathcal{I}$. This, in particular, holds for $C = A'$. For the induction step we have two cases. First, $H = H' \sqcap H''$. Due to $(\mathcal{I}', a) \models H'$ and $(\mathcal{I}', a) \models H''$, which is true by hypothesis, we immediately have that $(\mathcal{I}', a) \models H' \sqcap H''$. Second, $H = \exists r.H'$ for some role name r. Since we have $(\mathcal{I}, a) \models \exists r.H'$, there is an element a' with $(a, a') \in r^{\mathcal{I}}$ and hence, because a' is reachable from a in $\mathcal{I}$, we have that $(a, a') \in r^{\mathcal{I}'}$, and by induction hypothesis we have that $(\mathcal{I}', a') \models H'$, hence we have that $(\mathcal{I}', a) \models H$.

Since $\tau_{\textsf{exemplify-A}}$ exactly induces the substructure of a given input model with a and all its reachable elements, we can apply Lemma 8 to $\tau_{\textsf{exemplify-A}}(\mathcal{I}_\mathcal{T})$ and have that the image is a model of $\mathrm{S}_\mathcal{T}(A)$ because $\mathcal{T}$ entails $\mathrm{S}_\mathcal{T}(A)$.

In order to achieve minimality of the resulting structure through the model transformation as required in Definition 6, we compose $\tau_{\textsf{exemplify-A}}$ with a transduction $\tau_{\textsf{A-min}}$ that takes as an input a finite interpretation and yields a respective minimal part of a counter model. The idea of $\tau_{\textsf{A-min}}$ is the following. After the application of $\tau_{\textsf{exemplify-A}}$ to the canonical model $\mathcal{I}_\mathcal{T}$ of the TBox $\mathcal{T}$, the obtained substructure of $\mathcal{I}_\mathcal{T}$ satisfies the relevant implication sets, but is not minimal yet. This part we call *coarse relevant part of* $\mathcal{I}_\mathcal{T}$ w.r.t. the subsumption query and the TBox under consideration.

Transduction $\tau_{\textsf{A-min}}$ itself is a composition of two transductions. In the first step, superfluous roles of the coarse relevant part of $\mathcal{I}_\mathcal{T}$ are deleted, s.t. the model property of the coarse relevant part w.r.t. the relevant implication set is preserved. In the second step of $\tau_{\textsf{A-min}}$, the reachable part from the representative of concept A is cut out, for which the concept names and role names are induced much like in $\tau_{\textsf{exemplify-A}}$. This second transformation step of extracting the exemplify-A relevant part of a $\mathcal{I}_\mathcal{T}$, we name $\tau_{\textsf{A-reach}}$ and define it to be $\tau_{\textsf{exemplify-A}}$. For canonical models $\mathcal{I}_\mathcal{T}$ for some $\mathcal{EL}$ TBox $\mathcal{T}$, we have that $\tau_{\textsf{A-min}}(\mathcal{I}_\mathcal{T})$ is unique up to isomorphism. Hence, we write $\tau_{\textsf{A-min}}(\mathcal{I}_\mathcal{T}) = {\mathcal{I}_\mathcal{T}}'$, where ${\mathcal{I}_\mathcal{T}}'$ refers to a random element from the set of isomorphic smallest substructures.

The first step of $\tau_{\textsf{A-min}}$ deletes only superfluous roles and there might be several such sets of superfluous roles, which, strictly speaking, yield different substructures. To be formally accurate, we take into consideration that different selections of sets of superfluous roles might yield sets of substructures. However, instead of using parameters for the first step in the transduction $\tau_{\textsf{A-min}}$, we introduce sets of superfluous pairs in every role that serve for constructing the definition scheme for $\tau_{\textsf{min}}$. Let $\mathcal{I}$ be a finite interpretation and let $\Sigma = \textsf{sig}(\mathcal{I})$. We define $S_r(\mathcal{I})$ as a maximal subset of $r^\mathcal{I}$, s.t. $(\mathcal{I} \setminus S_r(\mathcal{I}), a) \models H \Leftrightarrow (\mathcal{I}, a) \models H$ for all concepts $H \in \mathcal{EL}(\Sigma)$ (and respectively for b if $B^\mathcal{I} \neq \emptyset$).We remind the reader that we assume a and b to be the representatives of A and B. Furthermore, we call $\mathcal{S}(\mathcal{I}) := \dot{\bigcup}_{r\in\Sigma} S_r(\mathcal{I})$ a *set of superfluous roles*, which is the disjoint union of the sets S_r indexed by the respective role name, for all role names in Σ. We omit writing $\mathcal{I}$ in $S_r(\mathcal{I})$ whenever it is clear from the context.

Definition 9 (Transduction $\tau_{\textsf{min}}$). *Let $\mathcal{I}$ be a finite interpretation over signature Σ, let $A, B \in \Sigma_\mathsf{C}$, and let $\mathcal{S}$ be a set of superfluous roles. The transduction $\tau_{\textsf{min}}$ is the composition of the transductions induced by the definition scheme:*

$$\delta(\mathcal{W}, x) := \textsf{True}, \quad \theta_N(\mathcal{W}, x) := N(x), \quad \eta_r(\mathcal{W}, x, y) := r(x, y) \wedge \neg S_r(x, y).$$

We define $\tau_{\textsf{A-min}}$ to be the composition $\tau_{\textsf{min}} \circ \tau_{\textsf{A-reach}}$. Different possible sets of superfluous roles $\mathcal{S}$ yield only isomorphic substructures for canonical models. Therefore, we will not explicitly mention $\mathcal{S}$.

Lemma 10. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, let $\mathcal{I}_\mathcal{T}$ be the canonical model of $\mathcal{T}$, and let $\Sigma = \mathsf{sig}(\mathcal{I}_\mathcal{T})$. Furthermore, let $A, B \in \Sigma_\mathsf{C}$, and let $\tau_{\mathsf{exemplify\text{-}A}}(\mathcal{I}_\mathcal{T}) = \mathcal{I}'$. Then, $\tau_{\mathsf{A\text{-}min}}(\mathcal{I}')$ is a minimal non-empty substructure of $\mathcal{I}_\mathcal{T}$, s.t. $\tau_{\mathsf{A\text{-}min}}(\mathcal{I}') \models \mathrm{S}_\mathcal{T}(A)$.*

Proof. Since $\tau_{\mathsf{A\text{-}min}}$ is model preserving by definition and Lemma 8, we have that $\mathcal{I}$, s.t. $a \in A^{\tau_{\mathsf{A\text{-}min}}(\mathcal{I}')}$ and that $\tau_{\mathsf{A\text{-}min}}(\mathcal{I}) \models \mathrm{S}_\mathcal{T}(A)$. Second to show is that $\tau_{\mathsf{min}}(\mathcal{I}')$ is indeed minimal, i.e., that for every strict substructure $\mathcal{I}^*$ of $\tau_{\mathsf{A\text{-}min}}(\mathcal{I}')$ we have that $\mathcal{I}^* \not\models \mathrm{S}_\mathcal{T}(A)$. This, however, is immediately given the definition of $\mathcal{S}$, since it's components are maximal subsets S_r for each role name in the signature, s.t. $(\mathcal{I} \setminus S_r(\mathcal{I}), a) \models H \Leftrightarrow (\mathcal{I}, a) \models H$. Then, the transduction $\tau_{\mathsf{A\text{-}min}}$ cuts out the reachable part from $a \in A^\mathcal{I}$ and induces the substructure w.r.t. Σ by the composition with τ_{reach}. Applying Lemma 8 additionally again implies that $\tau_{\mathsf{A\text{-}min}}(\mathcal{I}')$ is a minimal model of $\mathrm{S}_\mathcal{T}(A)$, meaning that one cannot remove another element, role or concept label.

The lemmata from above are used to show soundness of the transformations of canonical models, meaning that the obtained structures are indeed the exemplify-A-relevant part of the respective canonical models w.r.t. the subsumption query and TBox under consideration.

Theorem 11 (Soundness for exemplify-A-Relevance). *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, let $\phi \coloneqq A \sqsubseteq_\mathcal{T} B$ a subsumption query, s.t. $\mathcal{T} \not\models \phi$, and let $\mathcal{I}_\mathcal{T}$ the canonical model of $\mathcal{T}$. Then, we have that $\tau_{\mathsf{A\text{-}min}}(\tau_{\mathsf{exemplify\text{-}A}}(\mathcal{I}_\mathcal{T}))$ is the exemplify-A-relevant part of $\mathcal{I}_\mathcal{T}$ w.r.t. ϕ and $\mathcal{T}$.*

Proof. The transduction $\tau_{\mathsf{exemplify\text{-}A}}$ is always defined for the canonical model of an $\mathcal{EL}$ TBox if $\mathcal{T} \not\models \phi$. Fist, we need an element $a \in \Delta^{\tau_{\mathsf{exemplify\text{-}A}}(\mathcal{I})}$, s.t. $a \in A^{\tau_{\mathsf{exemplify\text{-}A}}(\mathcal{I})}$ and $a \notin B^{\tau_\alpha(\mathcal{I})}$. The precondition χ ensures that the parameter u is mapped to these conditions accordingly. The transduction then collects all reachable elements from u by δ. Due to the relation formulae θ and η, $\tau_{\mathsf{exemplify\text{-}A}}$ actually yields an induced substructure of $\mathcal{I}$. By Lemma 8 we have that this substructure is a model of $\mathrm{S}_\mathcal{T}(A)$. Also, $\tau_{\mathsf{exemplify\text{-}A}}(\mathcal{I})$ is not a model of ϕ by χ. By Lemma 10 we have that $\tau_{\mathsf{A\text{-}min}}(\tau_{\mathsf{exemplify\text{-}A}}(\mathcal{I}))$ satisfies the minimality condition of Definition 6.

Theorem 11 concludes the work on extracting exemplify-A-relevant parts from canonical models for $\mathcal{EL}$ TBoxes. We now move to extracting exemplify-A&B-relevant parts using the very same techniques as for exemplify-A-relevant parts.

Recall that exemplify-A&B-relevant parts of counter models are, in principle, the same as exemplify-A-relevant parts with the difference that the exemplify-A&B-relevant part also features a representative of B with all its properties as formulated in the TBox. Hence, we can easily apply the same techniques and lemmata as in the previous subsection. In order to construct a definition scheme that induces an MSO transduction that yields coarse exemplify-A&B-relevant parts of counter models, we modify the formulae of definition scheme $\mathsf{D}_{\mathsf{exemplify\text{-}A}}$. We will also have to adapt the minimality transduction accordingly.

Definition 12 (Transduction $\tau_{\textsf{exemplify-A\&B}}$). *Let Σ be a signature s.t. $A, B \in \Sigma_{\mathsf{C}} \subset \Sigma$, and let $\mathcal{W} = \{u, v\}$. The definition scheme $\mathsf{D}_{\textsf{exemplify-A\&B}}$ inducing the transduction $\tau_{\textsf{exemplify-A\&B}}$ consists of the formulae:*

$$\chi(\mathcal{W}) := A(u) \wedge \neg B(u) \wedge B(v), \qquad \theta_N(\mathcal{W}, x) := N(x),$$
$$\delta(\mathcal{W}, x) := reach(u, x) \vee reach(v, x), \qquad \eta_r(\mathcal{W}, x, y) := r(x, y).$$

Similarly to exemplify-A-relevance, we have to compose $\tau_{\textsf{exemplify-A\&B}}$ with a minimizing transduction that is applied the substructure of the canonical model under consideration. For this purpose, we redefine $\tau_{\textsf{A\&B-reach}}$ to be $\tau_{\textsf{exemplify-A\&B}}$ because we now also have to take care of the representative of B and its reachable elements. Now, we define $\tau_{\textsf{A\&B-min}}$ to be $\tau_{\textsf{min}} \circ \tau_{\textsf{A\&B-reach}}$.

Theorem 13 (Soundness for exemplify-A&B-Relevance). *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, let $\phi := A \sqsubseteq_{\mathcal{T}} B$ a subsumption query, s.t. $\mathcal{T} \not\models \phi$, and let $\mathcal{I}_{\mathcal{T}}$ a canonical model of $\mathcal{T}$. Then, we have that $\tau_{\textsf{A\&B-min}}(\tau_{\textsf{exemplify-A\&B}}(\mathcal{I}_{\mathcal{T}}))$ is the exemplify-A&B-relevant part of $\mathcal{I}_{\mathcal{T}}$ w.r.t. ϕ and $\mathcal{T}$.*

Proof. The proof follows the same argumentation as in the Theorem 11. The only difference is that there now also is a representative for concept B, which ensured by the precondition χ. From both the A (and not B) and the B representatives, all reachable elements are being collected and used to produce the induced substructure to satisfy Definition 6 applying Lemma 8 and Lemma 10.

Extracting the diff-relevant part of a canonical model is more difficult than the previous two cases. Here it is required that the representative of concept A must not satisfy subsumers of A w.r.t. $\mathcal{T}$ that are not subsumers of B w.r.t. $\mathcal{T}$; whereas the representative of concept B has no such restrictions, and hence must satisfy all subsumers of B w.r.t. $\mathcal{T}$. We make use of additional auxiliary predicates to extract the difference-based relevance parts of counter models. The predicate *sam* is true for two sets of elements if they constitute paths made of the same roles in $\mathcal{I}$. We denote the occurrence of free second-order variables X in a formula φ by box brackets in $\varphi[X]$.

$$sam[X, Y, a, b] := \exists h : \varphi[h, X, Y, a, b] \wedge \psi[h, X, Y, a, b] \text{ , where}$$
$$\psi[h] := \forall x, y, z, w : h(x, y) \wedge h(z, w) \to \bigvee_{r \in \Sigma_{\mathsf{R}}} r(x, z) \wedge r(y, w), \text{ and}$$

$\varphi[h, X, Y, a, b]$ defines h as a surjective map from Y to X with $h(b, a)$. To express that there is a path from element a to element b over the elements in X using the *succ* relation, we use an MSO formula $path[X, a, b]$ defined in [7, Proposition 5.11]. We combine these formulae in the predicate *sim*:

$$sim(a, b, x, y) := \exists X, Y : sam[Y, X, a, b] \wedge path[X, a, x] \wedge path[Y, b, y].$$

Note that *sam* contains a quantification over a binary relation h and the definition schemes using this formula are not inducing MSO transductions but rather second-order logic transductions. However, since the definition scheme formulae are evaluated on finite interpretations, *sam* is still decidable and hence the induced transduction computable.

Definition 14 (Transduction τ_{diff}). *Let Σ be a signature s.t. $A, B \in \Sigma_{\mathsf{C}} \subset \Sigma$, and let $\mathcal{W} = \{u, v\}$. The definition scheme $\mathsf{D}_{\mathit{diff}}$ inducing the transduction τ_{diff} consists of the formulae:*

$$
\begin{aligned}
\chi(\mathcal{W}) &:= A(u) \wedge \neg B(u) \wedge B(v),\\
\delta(\mathcal{W}, x) &:= x = u \vee [reach(u, x) \wedge \exists y : sim(u, v, x, y)] \vee reach(v, x),\\
\theta_A(\mathcal{W}, x) &:= [reach(u, x) \wedge A(x) \wedge \exists y : sim(u, v, x, y) \wedge A(y)] \vee x = u \vee\\
&\quad\ [reach(v, x) \wedge A(x)],\\
\theta_N(\mathcal{W}, x) &:= [reach(u, x) \wedge N(x) \wedge \exists y : sim(u, v, x, y) \wedge N(y)] \vee\\
&\quad\ [reach(v, x) \wedge N(x)] \quad \text{for all } N \in \Sigma_{\mathsf{C}} \setminus \{A\},\\
\eta_r(\mathcal{W}, x, y) &:= [reach(u, x) \wedge r(x, y) \wedge \exists z, w : r(z, w) \wedge sim(u, v, x, z) \wedge\\
&\quad\ sim(u, v, y, w)] \vee [reach(v, x) \wedge r(x, y)].
\end{aligned}
$$

Lemma 15. *Let $\mathcal{I}$ be an interpretation, let $A, B \in \mathsf{sig}(\mathcal{I})$ be two concept names, let $a \in (A \setminus B)^{\mathcal{I}}$, and let $b \in B^{\mathcal{I}}$. Then, for all $\mathcal{EL}$ concept descriptions H, with $H \neq A$, we have that $(\tau_{\mathit{diff}}(\mathcal{I}), a) \models H$ implies $(\tau_{\mathit{diff}}(\mathcal{I}), b) \models H$.*

Proof. By induction on the length of H. For the induction base, we assume H is a concept name other than A. By χ, we have $a, b \in \Delta^{\tau_{\mathsf{diff}}(\mathcal{I})}$ and by $\theta_H(\mathcal{W}, x)$, we have that $[reach(u, x) \wedge H(x) \wedge \exists y : sim(u, v, x, y) \wedge H(y)]$, and hence, by $h(b, a)$, we have that $H(b)$. Assume the claim holds for $|H| = n$. For the induction step, we have two cases: (1) $H = H_1 \sqcap H_2$ and (2) $H = \exists r.H_1$. For (1), the induction hypothesis applies to both H_1 and H_2, and hence, $(\tau_{\mathsf{diff}}(\mathcal{I}), b) \models H_1$ and $(\tau_{\mathsf{diff}}(\mathcal{I}), b) \models H_2$, and thus, $(\tau_{\mathsf{diff}}(\mathcal{I}), b) \models H_1 \sqcap H_2$. For (2), the induction hypothesis applies to H_1. By $\delta(\mathcal{W}, x)$, we have that $\mathcal{I} \models reach(a, x) \wedge \exists y : sim(u, v, x, y)$. Hence, there are reachable elements y from b and an element z reachable from b with $(\tau_{\mathsf{diff}}(\mathcal{I}), y) \models H_1$ and $r(z, y)$ and hence $(\tau_{\mathsf{diff}}(\mathcal{I}), b) \models H$.

As a consequence of Lemma 15 we have the following statement.

Corollary 16. *Let $\mathcal{T}$ be a TBox, let $\mathcal{I}$ be a model of $\mathcal{T}$, and let $\phi := A \sqsubseteq_{\mathcal{T}} B$ be a subsumption query, s.t. $\mathcal{I} \not\models \phi$. Then, for all $H \in \mathsf{Sub}_{\mathcal{T}}(A) \cap \mathsf{Sub}_{\mathcal{T}}(B)$, we have that $(\tau_{\mathit{diff}}(\mathcal{I}), a) \models H$.*

Theorem 17 (Soundness for diff-Relevance). *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, let $\phi := A \sqsubseteq_{\mathcal{T}} B$ a subsumption query, s.t. $\mathcal{T} \not\models \phi$, and let $\mathcal{I}_{\mathcal{T}}$ a canonical model of $\mathcal{T}$. Then, $\tau_{\mathit{A\&B\text{-}min}}(\tau_{\mathit{diff}}(\mathcal{I}_{\mathcal{T}}))$ is the diff-relevant part of $\mathcal{I}_{\mathcal{T}}$ w.r.t. ϕ and $\mathcal{T}$.*

Proof. As in Theorem 11, the transduction τ_{diff} is always defined for the canonical model of an $\mathcal{EL}$ TBox if $\mathcal{T} \not\models \phi$. First, having elements $a \in (A \setminus B)^{\tau_{\mathsf{diff}}(\mathcal{I})}$ and $b \in B^{\tau_{\mathsf{diff}}(\mathcal{I})}$ is ensured by the precondition χ again. Hence, we also have that $\tau_{\mathsf{diff}}(\mathcal{I}) \not\models \phi$. We need to show that $\tau_{\mathsf{diff}}(\mathcal{I}) \models \mathrm{C}_{\mathcal{T}}(A, B) \cup \mathrm{S}_{\mathcal{T}}(B)$. That means, for all $C \in \mathsf{Sub}_{\mathcal{T}}(H)$, we have to show $(\tau_{\mathsf{diff}}(\mathcal{I}), b) \models H$. This follows directly from Lemma 8, since τ_{diff} induces the substructure of b and all its reachable elements. We now have to show that $(\tau_{\mathsf{diff}}(\mathcal{I}), a) \models H$ for all $H \in \mathsf{Sub}_{\mathcal{T}}(G) \cap \mathsf{Sub}_{\mathcal{T}}(H)$, which follows from Corollary 16. Minimality follows the same argumentation as in Lemma 10.

Lastly, we define the transduction $\tau_{\textsf{flat-diff}}$ and prove its soundness for canonical models in a similar fashion as for the previous relevance parts.

Definition 18 (Transduction $\tau_{\textsf{flat-diff}}$). *Let Σ be a signature s.t. $A, B \in \Sigma_{\mathsf{C}} \subset \Sigma$, and let $\mathcal{W} = \{u, v\}$. The definition scheme $\mathsf{D}_{\textit{flat-diff}}$ inducing the transduction $\tau_{\textit{flat-diff}}$ consists of the formulae:*

$$
\begin{aligned}
\chi(\mathcal{W}) &:= A(u) \wedge \neg B(u) \wedge B(v), \\
\delta(\mathcal{W}, x) &:= [reach(u,x) \wedge \exists y :\ sim(u,v,x,y)] \vee \\
&\quad\ [reach(v,x) \wedge \exists y :\ sim(u,v,x,y)] \vee \\
&\quad\ [reach(v,x) \wedge \exists z : sim(u,v,x,z) \wedge succ(z,x)], \\
\theta_A(\mathcal{W}, x) &:= [reach(u,x) \wedge A(x) \wedge \exists y : sim(u,v,x,y) \wedge A(y)] \vee x = u \vee \\
&\quad\ [reach(v,x) \wedge A(x)], \\
\theta_N(\mathcal{W}, x) &:= [reach(u,x) \wedge N(x) \wedge \exists y : sim(u,v,x,y) \wedge N(y)] \vee \\
&\quad\ [reach(v,x) \wedge N(x)] \quad \textit{for all } N \in \Sigma_{\mathsf{C}} \setminus \{A\}, \\
\eta_r(\mathcal{W}, x, y) &:= [reach(u,x) \wedge r(x,y) \wedge \exists z, w : r(z,w) \wedge sim(u,v,x,z) \wedge \\
&\quad\ sim(u,v,y,w)] \vee [reach(v,x) \wedge r(x,y)].
\end{aligned}
$$

Theorem 19 (Soundness for $\textsf{flat-diff}$-Relevance). *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, let $\phi := A \sqsubseteq_{\mathcal{T}} B$ a subsumption query, s.t. $\mathcal{T} \not\models \phi$, and let $\mathcal{I}_{\mathcal{T}}$ a canonical model of $\mathcal{T}$. Then, $\tau_{\textit{A\&B-min}}(\tau_{\textit{flat-diff}}(\mathcal{I}_{\mathcal{T}}))$ is the $\textsf{flat-diff}$-relevant part of $\mathcal{I}_{\mathcal{T}}$ w.r.t. ϕ and $\mathcal{T}$.*

Proof. As before, the transduction $\tau_{\textsf{flat-diff}}$ is defined for $\mathcal{I}$ since $\mathcal{T} \not\models \phi$. Also, the precondition χ ensures together with θ_A and θ_B that $\tau_{\textsf{flat-diff}}(\mathcal{I}) \not\models \phi$. It also contains the A and B representatives by definition. As before, the substructure on the reachable elements from b is induced w.r.t. the signature of $\mathcal{I}$, and the concept and role labels for the by δ selected reachable elements from a are defined as in $\tau_{\textsf{diff}}$. Following the argumentation in Lemma 15 and Corollary 16, we show that $\tau_{\textsf{flat-diff}}(\mathcal{I}) \models \mathrm{C}_{\mathcal{T}}(A, B) \cup \bar{\mathrm{S}}_{\mathcal{T}}(A, B)$. To ensure $(\tau_{\textsf{flat-diff}}(\mathcal{I}), b) \models \bar{\mathrm{S}}_{\mathcal{T}}(A, B)$ we point to the disjunct $[reach(v,x) \wedge \exists z : sim(u,v,x,z) \wedge succ(z,x)]$ in δ. Since the δ also collects the direct successors of the leaves of the common part of the model starting in b compared to the part starting in a, we have satisfy (exactly) $\bar{\mathrm{S}}_{\mathcal{T}}(A, B)$. Furthermore, minimality stems from the arguments in Lemma 10.

Throughout the paper, we have assumed that the TBox is given in normal form. In order to get rid of the freshly introduced concept names, one can devise a definition scheme that simply induces the substructure of the model w.r.t. the signature of the not yet normalized TBox.

5 Conclusions and Future Work

We have introduced and motivated four notions of relevant parts of counter models for explaining non-subsumptions w.r.t. $\mathcal{EL}$ TBoxes, and we have devised sound means for extracting these parts by model transformation. We are currently implementing a system for providing explanations of $\mathcal{EL}$ non-subsumptions based on our relevance notions, to be evaluated on ontologies from practical applications. Possible extensions of this work are to consider DL knowledge bases that also contain data, as well as elaborating the methods for more expressive logics that also have the canonical model property, such as Horn DLs. In the long run, we would also like to consider different types of reasoning tasks such as explaining negative query answering results [6].

Acknowledgements This work was supported by the DFG in grant 389792660 as part of TRR 248 (https://perspicuous-computing.science), and QuantLA, GRK 1763 (https://lat.inf.tu-dresden.de/quantla).

References

1. Alrabbaa, C., Baader, F., Borgwardt, S., Koopmann, P., Kovtunova, A.: Finding small proofs for description logic entailments: Theory and practice. In: Proc. of the 23rd International Conference on Logic for Programming, Artificial Intelligence and Reasoning. EPiC Series in Computing, vol. 73, pp. 32–67. EasyChair (2020)
2. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proc. of the 19th International Joint Conference on Artificial Intelligence. pp. 364–369. Morgan-Kaufmann (2005)
3. Baader, F., Lutz, C., Suntisrivaraporn, B.: CEL — a polynomial-time reasoner for life science ontologies. In: Proc. of the 3rd International Joint Conference on Automated Reasoning. LNAI, vol. 4130, pp. 287–291. Springer-Verlag (2006)
4. Baader, F., Peñaloza, R., Suntisrivaraporn, B.: Pinpointing in the description logic $\mathcal{EL}^+$. In: Proc. of the 30th German Conference on Artificial Intelligence. LNCS, vol. 4667, pp. 52–67. Springer-Verlag (2007)
5. Brandt, S.: Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and — what else? In: Proc. of the 16th European Conference on Artificial Intelligence. pp. 298–302. IOS Press (2004)
6. Calvanese, D., Ortiz, M., Simkus, M., Stefanoni, G.: Reasoning about explanations for negative query answers in DL-lite. J. of Artif. Intel. Res. **48**, 635–669 (2013)
7. Courcelle, B., Engelfriet, J.: Graph Structure and Monadic Second-Order Logic — A Language-Theoretic Approach. Cambridge University Press (2012)
8. Dentler, K., Cornet, R., ten Teije, A., de Keizer, N.: Comparison of reasoners for large ontologies in the OWL 2 $\mathcal{EL}$ profile. Semantic Web **2**(2), 71–87 (2011)
9. Hieke, W., Turhan, A.: Towards model transformation in description logics — investigating the case of transductions. In: Proc. of the 6th Workshop on Formal and Cognitive Reasoning. CEUR, vol. 2680, pp. 69–82. CEUR-WS.org (2020)
10. Kazakov, Y., Krötzsch, M., Simancik, F.: The incredible ELK — from polynomial procedures to efficient reasoning with $\mathcal{EL}$ ontologies. J. of Autom. Reason. **53**(1), 1–61 (2014)
11. Lawley, M.J., Bousquet, C.: Fast classification in protégé: Snorocket as an OWL 2 EL reasoner. In: Proc. of the 6th Australasian Ontology Workshop. pp. 45–49. Australian Computer Society Inc. (2010)
12. Lee, Y.E.C., Williams, D.R., Anderson, J.F.: Frontal deficits differentiate progressive supranuclear palsy from parkinson's disease. J. of Neuropsy. **10**(1), 1–14 (2016)
13. Schlobach, S., Cornet, R.: Non-standard reasoning services for the debugging of description logic terminologies. In: Proc. of the 18th International Joint Conference on Artificial Intelligence. pp. 355–362. Morgan Kaufmann (2003)