Concept Definability and Interpolation in Enriched Models of $\mathcal{EL}$–TBoxes

Denis Ponomaryov[1] and Dmitry Vlasov[2]

[1] Institute of Artificial Intelligence, University of Ulm, Germany;
Institute of Informatics Systems, Novosibirsk, Russia
[2] Institute of Mathematics, Novosibirsk, Russia
ponom@iis.nsk.su, vlasov@academ.org

Abstract. It is known that in Description Logics explicit concept definability is directly related to concept interpolation. The problem to decide whether a concept is definable under a TBox wrt a signature usually reduces to entailment in the underlying logic. If an explicit definition exists, then it can be found as a concept interpolant for a concept inclusion entailed by an appropriately chosen TBox. In fact, it can be extracted from a corresponding proof of the concept inclusion. We describe a graph structure called enriched model that represents proofs in normalized $\mathcal{EL}$-TBoxes and show that, built once for a normalization of a given TBox $\mathcal{T}$, it can be used for deciding the existence or direct computation of explicit definitions of concepts under arbitrary subsets of axioms of $\mathcal{T}$ and wrt arbitrary subsignatures of $\mathcal{T}$. Solving this computational problem has applications in collaborative ontology engineering and is an important part of the recently proposed algorithms for ontology decomposition.

1 Introduction and Motivation

In the paper, we consider the Description Logic $\mathcal{EL}$ with the standard set of constructors for concepts, i.e. the conjunction, existential restriction, and the distinguished concept symbol $\top$. Consider a scenario in which several experts are working on the same ontology $\mathcal{T}$ and each expert has a vocabulary (a set of concept and role names) she is using for defining complex concepts in the ontology. Vocabularies need not be distinct between experts. If at some point an expert encounters an axiom in the ontology and a concept which is built using foreign vocabularies, she wants to figure out whether it still can be defined in her own vocabulary. Or at least, whether the encountered concept can be partially reformulated to use less symbols from foreign vocabularies (and more symbols from her own one). If the answer is "yes", then the expert has an option to see the actual reformulation of the concept and can replace it with the obtained reformulation (if she is allowed to do so under a collaborative ontology change policy). To illustrate this idea, consider the following simple example. Consider the TBox $\mathcal{T} = \{\varphi,\ A \sqsubseteq B,\ D \sqsubseteq A\}$, where $\varphi = A \sqcap B \sqsubseteq D$. Suppose an expert is interested in a reformulation of the concept $A \sqcap B$. Independently of what the expert's vocabulary is, this concept can be reformulated as A, since $\mathcal{T} \models A \sqcap B \equiv A$, and it can be replaced with A giving a TBox equivalent to $\mathcal{T}$. Now assume that the expert's vocabulary is $\{D\}$. We have $\mathcal{T} \models A \sqcap B \equiv D$, but if we replace the concept $A \sqcap B$ with D, we obtain the TBox which is not equivalent to the initial one. In other words, even though the required reformulation of the concept $A \sqcap B$ exists, it can not be used instead of $A \sqcap B$ without further modifications of the TBox (e.g. adding the inclusion $A \sqsubseteq D$ as an additional axiom to $\mathcal{T}$). The reason is that $\mathcal{T} \setminus \{\varphi\} \not\models A \sqcap B \equiv D$, i.e. the axiom φ is essential for proving this equivalence. Thus, the given example demonstrates the need to check for definability of concepts under certain subsets of a TBox. Now consider the TBox $\mathcal{S} = \{\varphi,\ A \sqsubseteq D,\ \exists r.D \sqsubseteq \exists r.A\}$, where $\varphi = \exists r.A \sqsubseteq B$, and assume that the expert's vocabulary is $\{D\}$. We have $\mathcal{S} \setminus \{\varphi\} \models \exists r.A \equiv \exists r.D$, i.e. the concept $\exists r.A$ can be partially reformulated as $\exists r.D$. Note that the subsignature which can not be changed under any reformulation wrt the expert's vocabulary (here, the single symbol r) may not be known in advance and has to be guessed. It is observed in the results of [6] however, that the mentioned problem of partial concept reformulation is not harder than entailment in the underlying logic.

Given a TBox $\mathcal{T}$, a concept C is said to be *explicitly definable* under $\mathcal{T}$ wrt a signature Δ if $\mathcal{T} \models C \equiv E$, where E is a concept over Δ. The problem to decide whether C is equivalent in $\mathcal{T}$ to a concept containing less symbols from $\mathtt{sig}\,(C) \setminus \Delta$ (thus, possibly having more Δ-symbols than C does) is called the problem of *partial concept reformulation* (under $\mathcal{T}$ wrt Δ). Explicit definability was studied in Description Logics in the context of query rewriting (Sect. 3 in [8]), acyclic representation of TBoxes [2] and is intimately related to modularity properties of ontologies via the notion of uniform interpolation. It is known that the Description Logic $\mathcal{EL}$ enjoys the concept interpolation property in the following sense (cf. Definition 3.1 in [3]): if the signatures of some TBoxes $\mathcal{T}_1, \mathcal{T}_2$ intersect by a set Δ, as well as signatures of some concepts C_1, C_2, then $\mathcal{T}_1 \cup \mathcal{T}_2 \models C_1 \sqsubseteq C_2$ yields the existence of a concept D in the signature Δ such that $\mathcal{T}_1 \cup \mathcal{T}_2 \models \{C_1 \sqsubseteq D,\ D \sqsubseteq C_2\}$ holds. It follows from this property that a concept C is explicitly definable under a TBox $\mathcal{T}$ wrt a subsignature $\Delta \subseteq \mathtt{sig}\,(\mathcal{T})$ iff $\mathcal{T} \cup \mathcal{T}^* \models C^* \sqsubseteq C$, where $\mathtt{sig}\,(\mathcal{T}) \cap \mathtt{sig}\,(\mathcal{T}^*) = \Delta$, $\mathtt{sig}\,(C^*) \cap \mathtt{sig}\,(C) \subseteq \Delta$, and $\mathcal{T}^*$, C^* are "copies" obtained from $\mathcal{T}$ and C, respectively, by replacing all non-Δ-symbols simultaneously with "fresh" ones, not present in $\mathtt{sig}\,(C) \cup \mathtt{sig}\,(\mathcal{T})$. Thus, deciding concept definability is polynomially reduced to entailment in $\mathcal{EL}$, although the explicit definition itself can be of exponential length in the size of $\mathcal{T}$ and C (cf. Example 28 in [6]). Note that the above construction of copies for $\mathcal{T}$ and C assumes that the signature Δ is known and it allows to check whether C is equivalent to a concept having symbols only from Δ (not necessarily all of them). In fact, the same idea can be used to check whether C can be partially reformulated wrt Δ. For instance, the concept $\exists r.A$ from the TBox S above allows for the partial reformulation wrt $\Delta = \{D\}$ and the entailment $(\mathcal{T} \setminus \{\varphi\}) \cup (\mathcal{T}^* \setminus \{\varphi^*\}) \models \exists r.A^* \sqsubseteq \exists r.A$ certifies this, where $\varphi^* = \exists r.A^* \sqsubseteq B$ and $\mathcal{T}^*$ and $\exists r.A^*$ are copies constructed as described above, with $\mathtt{sig}\,(\mathcal{T}) \cap \mathtt{sig}\,(\mathcal{T}^*) = \Delta \cup \{r\}$. Observe that we had to extend the initial signature Δ with the symbol r. In fact, if we extend an initial signature Δ with the set $\mathtt{sig}\,(C) \setminus \{x\}$ for a symbol $x \in \mathtt{sig}\,(C) \setminus \Delta$, then the entailment as above holds iff x can be removed in some partial reformulation of the concept C wrt Δ. Thus, in order to check for partial reformulations of concepts under a given TBox $\mathcal{T}$

wrt different signatures $\Delta \subseteq \mathtt{sig}(\mathcal{T})$, one needs to have an efficient procedure for checking the entailment

$$(\mathcal{T} \cup f^*(\mathcal{T})) \setminus \mathcal{S} \models f^*(C) \sqsubseteq C \qquad (*)$$

for subsets $\mathcal{S} \subseteq \mathcal{T} \cup f^*(\mathcal{T})$, different concepts C, and different functions f^* renaming signature symbols into fresh ones injectively and being the identity function on Δ. If the entailment holds, then the corresponding explicit definition of the concept C can be extracted from the proof of the inclusion $f^*(C) \sqsubseteq C$ from the TBox above. We will consider the case of this entailment problem, where C is a concept from the LHS or RHS of an axiom of $\mathcal{T}$, although the constructions proposed for solving the problem can be easily extended to the case of arbitrary concepts. Since the concept C, the set S, and function f^* are the changing parameters in the problem $(*)$, but the TBox $\mathcal{T}$ and the initial signature Δ are fixed, it makes sense to use a precomputed extended representation of $\mathcal{T}$ to be able to solve numerous instances of this problem efficiently.

The aim of the paper is to present the idea that the standard (canonical) model of an $\mathcal{EL}$-TBox with the closure operators S and R (in the sense of [1]) can be extended to become a structure (called *enriched model*) which, built once for a TBox $\mathcal{T}$, can be used as a precomputed template for solving problem $(*)$ with the changing input parameters (C, S, and f^*). An enriched model is an extended representation of a TBox used in the implementation of the decomposition algorithm for $\mathcal{EL}$–ontologies [6] which relies on solving numerous instances of the partial concept reformulation problem.

Our presentation is divided in three parts. In Section 3, we introduce the notion of enriched model and show that it can be used for checking the entailment of concept inclusions from arbitrary subsets of a TBox. In Section 4, we describe the process of computing explicit definitions of concepts in enriched models, which is conceptually similar to the idea of computing concept interpolants from tableaux (Sect. 4 in [8] and Sect. 3 in [3]). These results constitute the solution for entailment problem $(*)$ wrt the changing parameters which is described in Section 5. We note that the main idea behind the enriched models is the observation applicable to any system with non-trivial search problems: do a preprocessing of information in order to handle numerous standard calls to it faster.

2 Basic Notations

Let us introduce some necessary notations used in the paper. Let $\mathsf{N_c}$ and $\mathsf{N_r}$ stand, as usual, for the countably infinite sets of concept and role names in the alphabet of the language $\mathcal{EL}$. Denote $\mathsf{N_c^\top} = \mathsf{N_c} \cup \{\top\}$ and let $\bowtie$ be the symbol for either $\sqsubseteq$ or $\equiv$. A formula is an expression of the form $C_1 \bowtie C_2$, where C_1, C_2 are concepts, and a TBox is a finite set of formulas which does not contain $C_1 \equiv C_2$ and $C_1 \sqsubseteq C_2$ simultaneously for some concepts C_1, C_2. If Δ is a signature, then $Term(\Delta)$ will stand for the set of all concepts in the signature Δ and if $\Delta \subseteq \mathsf{N_c}$, then $Taut(\Delta)$ will denote the set of tautologies of the form $\top \sqsubseteq \top$, $A \sqsubseteq \top$, $A \sqsubseteq A$, for each $A \in \Delta$. Let $\mathcal{T}$ be a TBox. The notations $\mathtt{sub}(\mathcal{T})$, $\mathsf{N_c}^{\mathcal{T}}$, and $\mathsf{N_r}^{\mathcal{T}}$ will stand for the sets of all concepts, concept names, and role names respectively, occurring in axioms of $\mathcal{T}$, while $\mathsf{N_c^\top}(\mathcal{T})$ will be the shorthand for ${\mathsf{N_c}}^{\mathcal{T}} \cup \{\top\}$. The notation $\mathtt{sig}(\mathcal{T})$ will mean the signature of $\mathcal{T}$ and will be slightly abused to denote the signature of a formula or a concept. If $C \in \mathtt{sub}(\mathcal{T})$ is a concept and $A \in \mathsf{N_c}$, then $\mathcal{T}(C/A)$ will stand for the TBox obtained from $\mathcal{T}$ by substituting each occurrence of C in the LHS or RHS of axioms of $\mathcal{T}$ with the concept name A. We call a Tbox $\mathcal{T}'$ conservative extension of $\mathcal{T}$ if $\mathtt{sig}(\mathcal{T}) \subseteq \mathtt{sig}(\mathcal{T}')$ and for any formula φ, with $\mathtt{sig}(\varphi) \subseteq \mathtt{sig}(\mathcal{T})$, we have $\mathcal{T} \models \varphi$ iff $\mathcal{T}' \models \varphi$.

3 Proofs in Enriched Models of $\mathcal{EL}$-TBoxes

Informally, an enriched model is a graph structure for representing proofs in a deductive system for normal TBoxes. It is built iteratively in accordance with the inference rules and normal axioms obtained from an input TBox by normalization. In addition, an enriched model stores a map from normal to the original axioms to allow for checking the entailment of concept inclusions from different subsets of the input TBox. To illustrate this idea briefly, let us consider an example of eliminating a redundant axiom from a TBox.

Example 1 *Consider the TBox $\mathcal{S} = \{\varphi,\ A \sqsubseteq B,\ A \sqsubseteq D\}$, where $\varphi = A \sqcap B \sqsubseteq D$ and $A, B, D \in \mathsf{N_c}$. Note that either the first or the last axiom can be eliminated from $\mathcal{S}$ giving an equivalent TBox.*

Let $\mathcal{T} = \{\varphi', A \sqsubseteq B, A \sqsubseteq D,\ A \sqcap B \sqsubseteq X,\ X \sqsubseteq A, X \sqsubseteq B\}$ be the conservative extension of $\mathcal{S}$, where $\varphi' = X \sqsubseteq D$ and X is the "pseudonym" for the concept $A \sqcap B$. Note that $\mathcal{S} \setminus \{\varphi\} \models \varphi$ is equivalent to $\mathcal{T} \setminus \{\varphi'\} \models \varphi'$.

Consider the following graph, where every vertex has one or two labels depicted by the containment relation $\in f(.)$ and an arrow marked with δ or ρ. In the figure, if a vertex is marked with labels $f(X)$ and $\overset{\rho}{\mapsto} Y$, then $X, Y \in {\mathsf{N_c}}^{\mathcal{T}}$ and $\mathcal{T} \models Y \sqsubseteq X$. The vertices marked with δ correspond to axioms of $\mathcal{T}$ witnessing these entailments. Essentially, the graph represents possible proofs of primitive concept inclusions inside $\mathcal{T}$, note the two disconnected parts of the graph depicting independent proofs:

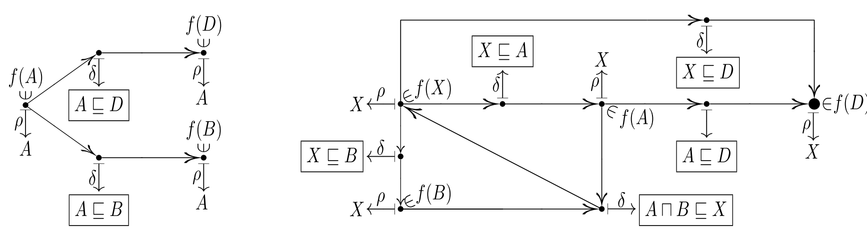

Figure 1. Example of enriched model: proofs as paths in the graph

The right-most vertex of the second graph component corresponds to the entailed concept inclusion $\varphi' = X \sqsubseteq D$. Observe that there is a subgraph which witnesses this entailment and does not contain vertices labelled by φ'. This corresponds to the fact that φ' is entailed by $\mathcal{T} \setminus \{\varphi'\}$ and thus, φ is redundant in $\mathcal{S}$.

The graph considered in the illustration generally corresponds to what we call further *enriched model* of a TBox. The rest of the section is devoted to a formal justification of the idea of checking for entailment of concept inclusions from subsets of a TBox by means of such structures.

Definition 1 (Primitivization). *A formula is called primitive if it is of the form $A \sqsubseteq B$ for $A, B \in \mathsf{N}_\mathsf{c}^\top$. A TBox $\mathcal{T}_p$ is called primitivization of a Tbox $\mathcal{T}$ if every axiom φ of $\mathcal{T}$ is equivalent in $\mathcal{T}_p$ to a primitive axiom $\varphi_p \in \mathcal{T}_p$, i.e. if $(\mathcal{T}_p \setminus \{\varphi_p\}) \cup \{\varphi\} \models \varphi_p$ and $\mathcal{T}_p \models \varphi$.*

An extended primitivization of a TBox $\mathcal{T}$ is a pair $\langle \mathcal{T}_p, \tau' \rangle$, where $\mathcal{T}_p$ is a primitivization of $\mathcal{T}$ and $\tau' : \mathcal{T}_p \to \mathcal{T} \cup \{null\}$ is a map, which are both obtained after the application of the following rewriting rule to each axiom of $\mathcal{T}$:

- *if $C_1 \bowtie C_2 \in \mathcal{T}$ then $\mathcal{T} := \mathcal{T}(C_i/A_i) \cup \{A_i \equiv C_i\}$, for each $C_i \notin \mathsf{N}_\mathsf{c}^\top$, $i = 1, 2$, where A_i is a fresh concept name,*
- *$\tau'(A_i \equiv C_i) = null$ and if φ_p is the image of $C_1 \bowtie C_2$ under the applied substitutions, then $\tau'(\varphi_p) = C_1 \bowtie C_2$.*

In other words, primitive formulas from $\mathcal{T}_p$ are in one-to-one correspondence with axioms of $\mathcal{T}$ which is represented by the map τ'. We note that in case $\varphi_p = C_1 \bowtie C_2$, i.e. when $C_1 \bowtie C_2$ is a primitive axiom of $\mathcal{T}$, we have $\tau'(C_1 \bowtie C_2) = C_1 \bowtie C_2$.

Example 2 *Let $\mathcal{T} = \{A_1 \sqsubseteq \exists r.B,\ \exists r.B \equiv A_2\}$, where $A_1, A_2 \in \mathsf{N}_\mathsf{c}$. Then the pair $\mathcal{T}_{ep} = \langle \{A_1 \sqsubseteq A, A \equiv A_2,\ A \equiv \exists r.B\},\ \tau' \rangle$ is an extended primitivization of $\mathcal{T}$, where $A \in \mathsf{N}_\mathsf{c}$ and $\tau'(A_1 \sqsubseteq A)$ equals $A_1 \sqsubseteq \exists r.B$, $\tau'(A \equiv A_2)$ is $\exists r.B \equiv A_2$, and $\tau'(A \equiv \exists r.B) = null$.*

Definition 2 (Normalization). *A formula is normal if it has one of the following forms: $A \sqsubseteq \exists r.B$, $\exists r.B \sqsubseteq A$, $A_1 \sqcap ... \sqcap A_n \sqsubseteq B$, for $n \geqslant 1$, with all A, B, A_i, $i = 1, .., n$, being concept names.A TBox is normal if each of its axiom is. A TBox $\mathcal{T}^n$ is called normalization of a TBox $\mathcal{T}$ if it is a conservative extension of $\mathcal{T}$ and each axiom of $\mathcal{T}^n$ is normal. Let $\mathcal{T}_{ep} = \langle \mathcal{T}_p, \tau' \rangle$ be an extended primitivization of a TBox $\mathcal{T}$. An extended normalization of $\mathcal{T}_{ep}$ is a pair $\langle \mathcal{T}^n, \tau \rangle$, where $\mathcal{T}^n$ is a normalization of $\mathcal{T}_p$ and $\tau : \mathcal{T}^n \to \mathcal{T}$ is a map such that:*

- *if $A \sqsubseteq B \in \mathcal{T}_p$, $A, B \in \mathsf{N}_\mathsf{c}^\top$, then $A \sqsubseteq B \in \mathcal{T}^{nm}$ and $\tau(A \sqsubseteq B) = \tau'(A \sqsubseteq B)$;*
- *if $A \equiv B \in \mathcal{T}_p$, $A, B \in \mathsf{N}_\mathsf{c}^\top$, then $\{A \sqsubseteq B,\ B \sqsubseteq A\} \in \mathcal{T}^{nm}$ and $\tau(A \sqsubseteq B) = \tau(B \sqsubseteq A) = \tau'(A \equiv B)$;*
- *$\tau(\varphi) = null$ for the rest of the axioms $\varphi \in \mathcal{T}^{nm}$.*

Example 3 *Let $\mathcal{T}_{ep}$ be the extended primitivization considered in Example 2. Then the extended normalization of $\mathcal{T}_{ep}$ is the pair $\mathcal{T}^{en} = \langle \{A_1 \sqsubseteq A,\ A \sqsubseteq A_2, A_2 \sqsubseteq A,\ A \sqsubseteq \exists r.B, \exists r.B \sqsubseteq A\}, \tau \rangle$, where $\tau(A_1 \sqsubseteq A)$ equals $A_1 \sqsubseteq \exists r.B$, $\tau(A \sqsubseteq A_2) = \tau(A_2 \sqsubseteq A) = \exists r.B \equiv A_2$, and $\tau(\varphi) = null$ for the remaining two axioms of $\mathcal{T}^{en}$.*

It follows from the standard decision procedure for $\mathcal{EL}$ (see e.g. [1]) that the following set of inference rules (under the notion of inference defined below) is complete wrt the semantic entailment of primitive formulas $A \sqsubseteq B$ from normal TBoxes (in the rules, $A, B, A_1, ..., A_n$, $n \geqslant 2$, stand for symbols from $\mathsf{N}_\mathsf{c}^\top$ and C is either a symbol from $\mathsf{N}_\mathsf{c}^\top$ or an expression $\exists r.X$, with $X \in \mathsf{N}_\mathsf{c}^\top$):

$$\text{(Ax)}\ \frac{}{A \sqsubseteq A} \qquad \text{(AxTop)}\ \frac{}{A \sqsubseteq \top} \qquad \text{(Trans)}\ \frac{A \sqsubseteq B}{A \sqsubseteq C}\ \ B \sqsubseteq C \in \mathcal{T}$$

$$\text{(AndR)}\ \frac{A \sqsubseteq A_1, ..., A \sqsubseteq A_n}{A \sqsubseteq B}\ \ A_1 \sqcap ... \sqcap A_n \sqsubseteq B \in \mathcal{T}$$

$$\text{(Ex)}\ \frac{A \sqsubseteq \exists r.A_1,\ A_1 \sqsubseteq A_2}{A \sqsubseteq B}\ \ \exists r.A_2 \sqsubseteq B \in \mathcal{T}$$

Figure 2. A deductive system for entailment of normal formulas

If $\mathcal{T}$ is a normal TBox and φ is a normal formula, then *an inference (a proof)* of φ from $\mathcal{T}$ is a tree, where φ is the root and every node is either a tautology $A \sqsubseteq A$, $A \sqsubseteq \top$, for $A \in \mathsf{N}_\mathsf{c}^\top$, or a formula obtained from the child nodes by one of the rules $Trans$, *AndR*, or Ex with the formulas in the side conditions being axioms of $\mathcal{T}$. If additionally, these formulas are from a subset $\mathcal{T} \setminus \mathcal{S}$, for some $\mathcal{S} \subseteq \mathcal{T}$, then we say that the inference *omits* $\mathcal{S}$.

We now introduce the notion of *enriched model* of a normal TBox. The definition consists of two parts: the description of an enriched model as an algebraic structure and the description of the process of its construction.

Let $\mathcal{T}$ be a normal TBox. An enriched model of $\mathcal{T}$ is a structure $\mathcal{M}^* = ((\Gamma^+, \Gamma^-, E), \delta, \rho, f)$, where (Γ^+, Γ^-, E) is a directed bipartite graph with the set E of edges and the set $\Gamma^+ \cup \Gamma^-$ of vertices augmented with maps:

- $\delta : \Gamma^- \to \mathcal{T} \cup Taut(\mathsf{N}_\mathsf{c}{}^\mathcal{T})$,
- $\rho : \Gamma^+ \to \mathsf{N}_\mathsf{c}^\top(\mathcal{T})\ \cup\ (\mathsf{N}_\mathsf{c}^\top(\mathcal{T}) \times \mathsf{N}_\mathsf{c}^\top(\mathcal{T}))$,
- $f : \mathsf{N}_\mathsf{c}^\top(\mathcal{T}) \cup \mathsf{N}_\mathsf{r}{}^\mathcal{T} \to 2^{\Gamma^+}$

such that the union of sets in the image of f is exactly Γ^+ and for all distinct $x, y \in \mathsf{N}_\mathsf{c}^\top(\mathcal{T}) \cup \mathsf{N}_\mathsf{r}{}^\mathcal{T}$ it holds that $f(x) \cap f(y) = \varnothing$.

By using the last restriction, we will sometimes abuse the notation and write $f^{-1}(v)$ to denote the preimage of the set $V \subseteq \Gamma^+$ with $v \in V$.

Informally, the plus-vertices of the graph (i.e. the elements of Γ^+) will correspond to concept and role names and will be mapped by ρ to concept names and pairs of concept names, respectively. The minus-vertices will correspond to axioms of $\mathcal{T}$ and tautologies. An enriched model $\mathcal{M}^* = ((\Gamma^+, \Gamma^-, E), \delta, \rho, f)$ of a normal TBox $\mathcal{T}$ is built iteratively in accordance with a set of graph expansion rules which correspond to the inference rules from Figure 2. For every $A \in \mathsf{N}_\mathsf{c}{}^\mathcal{T}$, the graph is first initialized with connected pairs of vertices $\langle w_1, y_1 \rangle$ and $\langle w_2, y_2 \rangle$ with $w_1, w_2 \in \Gamma^-$ and $y_1, y_2 \in \Gamma^+$, such that $y_1 \in f(A)$, $\rho(y_1) = A$, $\delta(w_1)$ is a tautology of the form $A \sqsubseteq A$ and $y_2 \in f(\top)$, $\rho(y_2) = A$, and $\delta(w_2)$ is $A \sqsubseteq \top$. Besides, a pair $\langle w, y \rangle$ is initialized, with $y \in f(\top)$, $\rho(y) = \top$, and $\delta(w) = \top \sqsubseteq \top$. This is the initialization step for the bipartite graph and the mappings δ, ρ, f. To formulate the iterative construction process of $\mathcal{M}^*$, we present it as a set of graph expansion rules, which describe how the bipartite graph should be expanded (and the mappings updated) based on the previously built vertices and an axiom from $\mathcal{T}$ encountered at an iteration step. Each rule has the form ($m \geqslant 2$):

$$\frac{\bullet(U_1, V_1), ..., \bullet(U_{m-1}, V_{m-1})}{\bullet(U_m, V_m)} \quad \varphi$$

where for $i = 1, .., m$, $U_i \in \mathsf{N}_\mathsf{c}^\top(\mathcal{T}) \cup \mathsf{N}_\mathsf{r}{}^\mathcal{T}$ and V_i is either a symbol from $\mathsf{N}_\mathsf{c}^\top(\mathcal{T})$ or a pair of symbols from $\mathsf{N}_\mathsf{c}^\top(\mathcal{T})$, and φ is an axiom of $\mathcal{T}$.
The rule is read as follows:

- if there is an axiom φ in $\mathcal{T}$
- and for each (U_i, V_i), $i = 1, .., m-1$, there is a vertex $x_i \in \Gamma^+$ such that $x_i \in f(V_i)$ and $\rho(x_i) = U_i$
- if there does not exist a vertex $y \in \Gamma^+$ such that $y \in f(V_m)$ and $\rho(y) = U_m$, then initialize such vertex and
- if there does not already exist a vertex $w \in \Gamma^-$, with $\delta(w) = \varphi$, such that $\langle x_i, w\rangle \in E$ for all $i = 1, .., m-1$ and $\langle w, y\rangle \in E$, then initialize such vertex.

The list of the graph expansion rules is presented in the figure below ($n \geqslant 2$ in the rule $AndR^\bullet$). Each rule is augmented with an illustration of the constructed part of the graph with the above mentioned vertices x_i, w, and y, $i = 1, .., m-1$.

$$(Trans_1^\bullet)\ \frac{\bullet(A,B)}{\bullet(A,C)}\ B \sqsubseteq C$$

$$(Trans_2^\bullet)\ \frac{\bullet(A,B)}{\bullet(r,\langle A,C\rangle)}\ B \sqsubseteq \exists r.C$$

$$(AndR^\bullet)\ \frac{\bullet(A,A_1),...,\bullet(A,A_n)}{\bullet(A,B)}\ A_1 \sqcap ... \sqcap A_n \sqsubseteq B$$

$$(Ex^\bullet)\ \frac{\bullet(r,\langle A,A_1\rangle),\ \bullet(A_1,A_2)}{\bullet(A,B)}\ \exists r.A_2 \sqsubseteq B$$

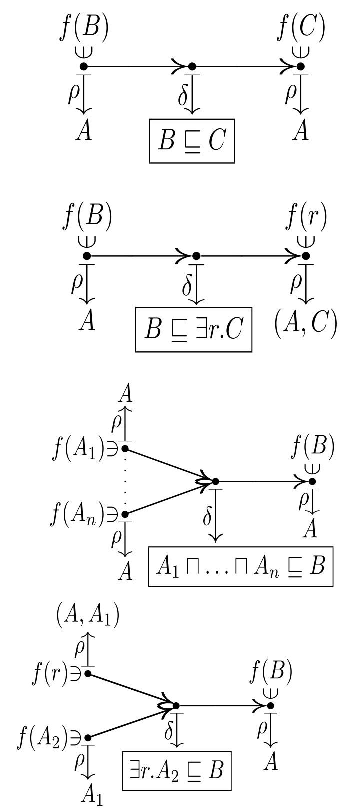

Figure 3. Iterative construction of an enriched model: graph expansion rules

The iterative procedure of model construction finishes when no rule results in an expansion of the graph, i.e. when no vertex or edge can be added.

Definition 3 (Enriched model of a normal TBox). *Let $\mathcal{T}$ be a normal TBox. The structure $\mathcal{M}^* = ((\Gamma^+, \Gamma^-, E), \delta, \rho, f)$ built for $\mathcal{T}$ as above is called enriched model of $\mathcal{T}$.*

Definition 4 (Submodel of an enriched model). *Let $\mathcal{M}^* = (\Gamma = (\Gamma^+, \Gamma^-, E), \delta, \rho, f)$ be an enriched model. The structure $\mathcal{M}^*_s = (\Gamma_s = (\Gamma^+_s, \Gamma^-_s, E_s), \delta_s, \rho_s, f_s)$ is called a submodel of $\mathcal{M}^*$ if Γ_s is a subgraph of Γ, each δ_s, ρ_s, f_s is the restriction of the corresponding map from $\mathcal{M}^*$ onto the vertices of Γ_s and the following holds:*

- *for every $v^- \in \Gamma^-_s$, all vertices $v^+ \in \Gamma^+$ with $(v^+, v^-) \in E$ are in Γ^+_s and connected with v^- in Γ_s;*
- *for every $v^+ \in \Gamma^+_s$, there exists $v^- \in \Gamma^-_s$ such that $(v^-, v^+) \in E_s$; if such v^- is unique, then the submodel is called deterministic.*

In general, the bipartite graph (Γ^+, Γ^-, E) may not have a tree-shape due to cycles caused e.g. by equalities between concept names (recall Example 1). However, there always exist submodels which are trees and these submodels are in one-to-one correspondence with proofs in the deductive system of Figure 2.

Lemma 1 (Deterministic tree submodels correspond to proofs) *Let $\mathcal{M}^* = (\Gamma = (\Gamma^+, \Gamma^-, E), \delta, \rho, f)$ be an enriched model of a TBox $\mathcal{T}$. For any vertex $v^+ \in \Gamma^+$, there exists a deterministic submodel $\mathcal{M}^*_s = (\Gamma_s = (\Gamma^+_s, \Gamma^-_s, E_s), \delta_s, \rho_s, f_s)$, where Γ_s is a tree with the root v^+ and each leaf is a vertex $v^- \in \Gamma^-$ with $\delta(v^-) \in Taut(\mathsf{N_c}^{\mathcal{T}})$. Let $\mathcal{S} = \mathcal{T} \setminus \{\delta(v^-) \mid v^- \in \Gamma^-_s\}$.*

The submodel $\mathcal{M}^*_s$ corresponds to an inference from $\mathcal{T}$ omitting $\mathcal{S}$ of the concept inclusion $\rho(v^+) \sqsubseteq f^{-1}(v^+)$ (in case $\rho(v^+) \in \mathsf{N_c}$) and of the inclusion $A \sqsubseteq \exists f^{-1}(v^+).B$ (in case $\rho(v^+) = \langle A, B\rangle$).

Proof. The claim is proved by induction on the number of iteration steps in the construction process of $\mathcal{M}^*$ for $\mathcal{T}$ which are needed to initialize the vertex v^+. The base case is trivial by the initialization step for $\mathcal{M}^*$: in this case the required tree is just a two-element chain corresponding to an inference of a tautology from $Taut(\mathsf{N_c}^{\mathcal{T}})$. This inference trivially omits $\mathcal{S}$. The induction step is simple as well, since at each iteration step of the graph expansion, by the induction hypothesis, for every plus-vertex in the premise of an expansion rule, there is a corresponding deterministic submodel, the part of the graph to be constructed has the form of a tree (Figure 2) and the intermediate vertex v^- to be constructed satisfies $\delta(v^-) \in \mathcal{T}$. If the expansion rule applied is $Trans^\bullet_i$ for some $i = 1, 2$, then the corresponding inference of the concept inclusion is by the rule $Trans$ with the axiom in the side condition equal to $\delta(v^-)$; the rest of the rules in Figure 3 are in direct correspondence to the rules from Figure 2. □

Let us call *proof submodel* a submodel with the properties from Lemma 1 for some plus-vertex v^+. As expected, the converse of the lemma holds as well.

Lemma 2 (Proofs correspond to deterministic tree submodels) *Let $\mathcal{T}$ be a normal TBox, $\mathcal{M}^*$ be an enriched model of $\mathcal{T}$, and let φ be a formula of either of the forms $A \sqsubseteq B$, $A \sqsubseteq \exists r.B$, for $A, B \in \mathsf{N_c}$. If there is an inference of φ from $\mathcal{T}$ which omits a subset $\mathcal{S} \subseteq \mathcal{T}$, then there is a proof submodel of $\mathcal{M}^*$ with the root vertex v^+ corresponding to φ and each non-leaf minus-vertex v^- satisfying $\delta(v^-) \in (\mathcal{T} \setminus \mathcal{S})$.*

Thus, in the notations of the lemma, there is an inference of φ from $\mathcal{T}$ iff there is a plus-vertex $v^+ \in \Gamma^+$ in the enriched model of $\mathcal{T}$ such that $v^+ \in f(B)$ and $\rho(v^+) = A$, or $v^+ \in f(r)$ and $\rho(v^+) = \langle A, B \rangle$, depending on the form of φ. By the construction of the enriched model, the vertex v^+ in the graph is unique for φ, so the number of plus-vertices in the model is at worst $O(n^2)$, where n in the size of the input TBox $\mathcal{T}$. Every path of length 3 in the graph starting and ending in a plus-vertex has a minus-vertex in the middle (the graph is bipartite). Every minus-vertex is labeled with an axiom of $\mathcal{T}$ and, by the construction, the number of minus-vertices between two plus-vertices does not exceed the number of axioms in $\mathcal{T}$ (there can not exist two such minus-vertices v_1, v_2 with $\delta(v_1) = \delta(v_2)$). Thus, the size of the enriched model is polynomial in the size of $\mathcal{T}$. Note that besides the information about all the entailed primitive concept inclusions, the enriched model contains information about all their possible inferences from axioms of $\mathcal{T}$ and tautologies $Taut(\mathsf{N_c}^{\mathcal{T}})$, recall also the graphical representation of independent inferences from Example 1.

Lemma 3 (Checking for entailment with a normal TBox) *Let $\mathcal{T}$ be a TBox, $\mathcal{S} \subseteq \mathcal{T}$ be a subset of axioms, and $\varphi \in \mathcal{T}$ be an axiom. Let $\mathcal{T}^{ep} = \langle \mathcal{T}_p, \tau' \rangle$ be an extended primitivization of $\mathcal{T}$ and $\mathcal{T}^{en} = \langle \mathcal{T}^{norm}, \tau \rangle$ be an extended normalization of $\mathcal{T}^{ep}$. Denote $\mathcal{S}^{norm} = \{\psi \in \mathcal{T}^{norm} \mid \tau(\psi) \in \mathcal{S}\}$ and $\varphi^{norm} = \{\psi \in \mathcal{T}^{norm} \mid \tau(\psi) = \varphi\}$. Then we have $\mathcal{T} \setminus \mathcal{S} \models \varphi$ iff $\mathcal{T}^{norm} \setminus \mathcal{S}^{norm} \models \varphi^{norm}$.*

Proof. For brevity we consider only the special case, when $\mathcal{S} = \{\varphi\}$ and φ is of the form $C_1 \equiv C_2$ for some concepts $C_1, C_2 \notin \mathsf{N_c^\top}$. It follows that the set φ^{norm} is not a singleton. Other cases, when $\mathcal{S}$ is an arbitrary subset of axioms, or φ is of the form $C_1 \sqsubseteq C_2$, or $C_i \in \mathsf{N_c^\top}$, for some $i = 1, 2$, are considered analogously. Let $\varphi_p = A_1 \equiv A_2$ be the primitive axiom of $\mathcal{T}_p$ such that $\tau'(\varphi_p) = \varphi$ and $A_i \equiv C_i \in \mathcal{T}_p$ for $i = 1, 2$. Then, by the definition of extended normalization, $\varphi^{norm} = \{A_1 \sqsubseteq A_2,\ A_2 \sqsubseteq A_1\}$ (with each of the concept inclusions mapped by τ to φ). Consider the pair $\langle (\mathcal{T} \setminus \{\varphi\})_p, \tilde{\tau}) \rangle$, where $(\mathcal{T} \setminus \{\varphi\})_p = \mathcal{T}_p \setminus (\{\varphi_p\} \cup \bigcup_{i=1,2}\{A_i \equiv C_i \mid C_i$ is not the LHS or RHS of an axiom of $\mathcal{T} \setminus \{\varphi\}\})$ and $\tilde{\tau}$ is the restriction of the map τ' onto the set $\mathcal{T}_p \setminus \{\varphi_p\}$. Clearly, $\langle (\mathcal{T} \setminus \{\varphi\})_p, \tilde{\tau}) \rangle$ is an enriched primitivization of $\mathcal{T} \setminus \{\varphi\}$ and hence, its conservative extension. Moreover, any model of $(\mathcal{T} \setminus \{\varphi\})_p$ expands to a model of $\mathcal{T}_p \setminus \{\varphi_p\}$ and thus, $\mathcal{T}_p \setminus \{\varphi_p\}$ is a conservative extension of $(\mathcal{T} \setminus \{\varphi\})_p$. Now consider the set $\mathcal{T}^{norm} \setminus \varphi^{norm}$. It is easy to see that this set is a normalization of $\mathcal{T}_p \setminus \{\varphi_p\}$ and thus, its conservative extension.

($\Rightarrow$) : If $\mathcal{T} \setminus \{\varphi\} \models \varphi$, then $(\mathcal{T} \setminus \{\varphi\})_p \models C_1 \equiv C_2$ and $\mathcal{T}_p \setminus \{\varphi_p\} \models \{C_1 \equiv C_2,\ A_1 \equiv C_2, A_2 \equiv C_2\}$. But then $\mathcal{T}^{norm} \setminus \varphi^{norm}$ entails this set of formulas too and hence, $\mathcal{T}^{norm} \setminus \varphi^{norm} \models \varphi^{norm}$.

($\Leftarrow$) : Assume that φ^{norm} is entailed by $\mathcal{T}^{norm} \setminus \varphi^{norm}$. Then it is entailed by $\mathcal{T}_p \setminus \{\varphi_p\}$ and, since $\mathcal{T}_p \setminus \{\varphi_p\} \models \{A_1 \equiv C_2, A_2 \equiv C_2\}$, we have $\mathcal{T}_p \setminus \{\varphi_p\} \models \varphi$. As

$\mathcal{T}_p \setminus \{\varphi_p\}$ is a conservative extension of $\mathcal{T} \setminus \{\varphi\}$, we conclude that $\mathcal{T} \setminus \{\varphi\} \models \varphi$. $\square$

Theorem 1 (Entailment from subset of TBox in enriched model). *Let $\mathcal{T}$ be a TBox and $\mathcal{T}^{ep} = \langle \mathcal{T}_p, \tau' \rangle$ be an extended primitivization of $\mathcal{T}$. Let $\langle \mathcal{T}^{norm}, \tau \rangle$ be an extended normalization of $\mathcal{T}^{ep}$ and $\mathcal{M}^*$ be an enriched model of $\mathcal{T}^{norm}$. Let C_1, C_2 be concepts appearing on the LHS or RHS of an axiom of $\mathcal{T}$ and for $i = 1, 2$, let $A_i \in \mathtt{sig}(\mathcal{T}_p)$ be a concept name which is either C_i (if $C_i \in \mathsf{N_c}$), or such that $A_i \in \mathtt{sig}(\mathcal{T}_p) \setminus \mathtt{sig}(\mathcal{T})$ and $A_i \equiv C_i \in \mathcal{T}_p$. Then, for a subset $\mathcal{S} \subseteq \mathcal{T}$ of axioms, it holds $\mathcal{T} \setminus \mathcal{S} \models C_1 \sqsubseteq C_2$ iff there is a proof submodel $\mathcal{M}_s = (\Gamma_s = (\Gamma_s^+, \Gamma_s^-, E_s), \delta_s, \rho_s, f_s)$ of $\mathcal{M}^*$ and a vertex $v^+ \in \Gamma_s^+$ such that:*

- *$v^+ \in f_s(A_2)$ and $\rho_s(v^+) = A_1$,*
- *for any non-leaf vertex $v^- \in \Gamma_s^-$, $\tau(\delta_s(v^-)) \notin \mathcal{S}$.*

Proof. Let us denote $\varphi = C_1 \sqsubseteq C_2$. If $\varphi \in \mathcal{T}$ then, by the definitions of extended primitivization and normalization, $A_1 \sqsubseteq A_2 \in \mathcal{T}^{norm}$, $\tau(A_1 \sqsubseteq A_2) = C_1 \sqsubseteq C_2$ and thus, by Lemma 3, the condition $\mathcal{T} \setminus \mathcal{S} \models C_1 \sqsubseteq C_2$ is equivalent to $\mathcal{T}^{norm} \setminus \mathcal{S}^{norm} \models A_1 \sqsubseteq A_2$, where $\mathcal{S}^{norm} = \{\psi \in \mathcal{T}^{norm} \mid \tau(\psi) \in \mathcal{S}\}$. If $\varphi \notin \mathcal{T}$ then consider the TBox $\mathcal{Q} = \mathcal{T} \cup \{\varphi\}$. Clearly, $\mathcal{Q}^{ep} = \langle \mathcal{T}_p \cup \{A_1 \sqsubseteq A_2\}, \tau'_{\mathcal{Q}} \rangle$ is an extended primitivization of $\mathcal{Q}$ and $\langle \mathcal{T}^{norm} \cup \{A_1 \sqsubseteq A_2\}, \tau_{\mathcal{Q}} \rangle$ is an extended normalization of $\mathcal{Q}^{ep}$, where $\tau_{\mathcal{Q}}(A_1 \sqsubseteq A_2) = \tau'_{\mathcal{Q}}(A_1 \sqsubseteq A_2) = \varphi$ and $\tau'_{\mathcal{Q}}$, $\tau_{\mathcal{Q}}$ are equal to τ and τ' respectively, on the rest of their domains $\mathcal{T}^{norm}$ and $\mathcal{T}_p$. Observe that $\mathcal{T} \setminus \mathcal{S} \models \varphi$ is the same as $\mathcal{Q} \setminus (\mathcal{S} \cup \{\varphi\}) \models \varphi$, which by Lemma 3 and the above noted is equivalent to $(\mathcal{T}^{norm} \cup \{A_1 \sqsubseteq A_2\}) \setminus \mathcal{P} \models A_1 \sqsubseteq A_2$, where $\mathcal{P} = \{\psi \in \mathcal{T}^{norm} \cup \{A_1 \sqsubseteq A_2\} \mid \tau(\psi) \in \mathcal{S} \cup \{\varphi\}\}$. We have $A_1 \sqsubseteq A_2 \in \mathcal{P}$, so the last entailment is the same as $\mathcal{T}^{norm} \setminus \mathcal{S}^{norm} \models A_1 \sqsubseteq A_2$. Therefore, in each of the two considered cases, the condition $\mathcal{T} \setminus \mathcal{S} \models C_1 \sqsubseteq C_2$ is equivalent to $\mathcal{T}^{norm} \setminus \mathcal{S}^{norm} \models A_1 \sqsubseteq A_2$.

$(\Rightarrow)$: If we have the entailment above, then there is an inference of $A_1 \sqsubseteq A_2$ from $\mathcal{T}^{norm}$ which omits $\mathcal{S}^{norm}$. Then, by Lemma 2, there are the required proof submodel $\mathcal{M}_s$ and the vertex $v^+ \in \Gamma_s^+$ such that for all non-leaf vertices $v^- \in \Gamma^-$ we have $\delta_s(v^-) \notin \mathcal{S}^{norm}$ and hence, $\tau(\delta_s(v^-)) \notin \mathcal{S}$.

$(\Leftarrow)$: Suppose there is a proof submodel $\mathcal{M}_s$ of $\mathcal{M}^*$ satisfying the conditions of the theorem. Then, by the definition of $\mathcal{S}^{norm}$ and by Lemma 1, there is an inference of $A_1 \sqsubseteq A_2$ from $\mathcal{T}^{norm}$ which omits $\mathcal{S}^{norm} \setminus Taut(\mathsf{N_c}^{\mathcal{T}})$, but then $\mathcal{T}^{norm} \setminus (\mathcal{S}^{norm} \setminus Taut(\mathsf{N_c}^{\mathcal{T}})) \models A_1 \sqsubseteq A_2$ and $(\mathcal{T}^{norm} \setminus \mathcal{S}^{norm}) \cup (\mathcal{T}^{norm} \cap Taut(\mathsf{N_c}^{\mathcal{T}})) \models A_1 \sqsubseteq A_2$, hence $\mathcal{T}^{norm} \setminus \mathcal{S}^{norm} \models A_1 \sqsubseteq A_2$ and we have $(\mathcal{T}^{norm} \cup \{A_1 \sqsubseteq A_2\}) \setminus \mathcal{P} \models A_1 \sqsubseteq A_2$. By Lemma 3, this entailment is equivalent to $\mathcal{Q} \setminus (\mathcal{S} \cup \{C_1 \sqsubseteq C_2\}) \models C_1 \sqsubseteq C_2$ which is the same as $\mathcal{T} \setminus \mathcal{S} \models C_1 \sqsubseteq C_2$. $\square$

Therefore, in order to check the entailment $\mathcal{T} \setminus \mathcal{S} \models C_1 \sqsubseteq C_2$, it suffices to find first a plus-vertex v^+ witnessing the entailment $\mathcal{T}^{norm} \models A_1 \sqsubseteq A_2$ in the enriched model $\mathcal{M}^*$ (the vertex must satisfy the conditions $v^+ \in f_s(A_2)$ and $\rho_s(v^+) = A_1$) and if it exists, find an inference of $A_1 \sqsubseteq A_2$, which does not refer via the map $\tau \circ \delta$ to formulas of $\mathcal{S}$, as a proof submodel of $\mathcal{M}^*$ containing v^+. The procedure of finding this submodel reduces to a traversal of the graph (Γ^+, Γ^-, E) of $\mathcal{M}^*$ starting from the vertex v^+. The procedure is conceptually similar to the one presented in the context of computing concept interpolants in Section 4 and is therefore skipped.

4 Computing Explicit Definitions with Enriched Models

Let $\mathcal{T}_1$ be a TBox, C_1 be a concept from the LHS or RHS of an axiom of $\mathcal{T}$, $\mathcal{S}_1 \subseteq \mathcal{T}_1$ be a subset of axioms, and Δ be a signature. Suppose we want to check whether C_1 is Δ–definable under $\mathcal{T}_1 \setminus \mathcal{S}_1$. Let $\mathcal{T}_2 = f^*(\mathcal{T}_1)$ and $\mathcal{S}_2 = f^*(\mathcal{S}_1) \subseteq \mathcal{T}_2$ be the corresponding "copy"-TBoxes (with $\mathtt{sig}\,(\mathcal{T}_1) \cap \mathtt{sig}\,(\mathcal{T}_2) = \Delta$) and let $C_2 = f^*(C_1)$ be the corresponding "copy"-concept obtained by a renaming function f^*. We may assume that $f^*(\mathcal{T}_1 \setminus \mathcal{S}_1) = \mathcal{T}_2 \setminus \mathcal{S}_2$ and thus, by concept interpolation, definability is equivalent to the condition $(\mathcal{T}_1 \setminus \mathcal{S}_1) \cup (\mathcal{T}_2 \setminus \mathcal{S}_2) \models C_1 \sqsubseteq C_2$. The observations from Section 3 imply that it suffices to consider the case when the TBox $\mathcal{T}_1$ is normal and C_i is a concept name from $\mathtt{sig}\,(\mathcal{T}_i) \setminus \Delta$, for $i = 1, 2$. Then $(\mathcal{T}_1 \setminus \mathcal{S}_1) \cup (\mathcal{T}_2 \setminus \mathcal{S}_2) \models C_1 \sqsubseteq C_2$ holds iff there is a submodel $\mathcal{M}_s$ in an enriched model $\mathcal{M}^* = ((\Gamma^+, \Gamma^-, E),\ \delta, \rho, f)$ of $\mathcal{T}_1 \cup \mathcal{T}_2$ and a vertex $v^+ \in \Gamma^+$ certifying this entailment. Let us show how a concept interpolant for the inclusion $C_1 \sqsubseteq C_2$ (i.e. an explicit Δ-definition of C_1 under $\mathcal{T} \setminus \mathcal{S}$) can be computed while searching for this submodel in $\mathcal{M}^*$. As Lemmas 1 and 2 evidence, the submodel $\mathcal{M}_s$ must be in direct correspondence with a proof of $C_1 \sqsubseteq C_2$ from $(\mathcal{T}_1 \setminus \mathcal{S}_1) \cup (\mathcal{T}_2 \setminus \mathcal{S}_2)$. We have $\mathtt{sig}\,(\mathcal{T}_1) \cap \mathtt{sig}\,(\mathcal{T}_2) = \Delta$ and $C_i \in \mathtt{sig}\,(\mathcal{T}_i) \setminus \Delta$, for $i = 1, 2$, so it suffices to understand how a concept interpolant between C_1 and C_2 can be extracted from a proof of $C_1 \sqsubseteq C_2$ in this situation, which is the purpose of Theorem 2. The property of concept interpolation shown in the theorem is well-known for $\mathcal{EL}$, but the proof is given to justify soundness of the procedure for interpolant computation described at the end of this section. First, let us consider the following example illustrating the main idea.

Example 4 *Consider the TBox $\mathcal{S} = \{A \sqcap B \sqsubseteq C,\ A \equiv D\}$, with $A, B, C, D \in \mathsf{N_c}$, the corresponding primitivization $\mathcal{T}_p = \{X \sqsubseteq C,\ A \equiv D, X \equiv A \sqcap B\}$, and a normalization $\mathcal{T}$ of $\mathcal{T}_p$. Denote $\Delta = \{D, B\}$ and let $\mathcal{T}^*$ be the "copy" of $\mathcal{T}$ under the injective renaming of all non-Δ-symbols into fresh ones such that X^* and C^* are the "copies" of X and C in $\mathcal{T}^*$, respectively. Then the entailment $\mathcal{S} \setminus \{A \sqcap B \sqsubseteq C\} \models A \sqcap B \equiv D \sqcap B$ is equivalent to $(\mathcal{T} \setminus \{X \sqsubseteq C\}) \cup (\mathcal{T}^* \setminus \{X^* \sqsubseteq C^*\}) \models X^* \sqsubseteq X$ and there is a vertex and a subgraph in the enriched model of $\mathcal{T} \cup \mathcal{T}^*$ certifying this entailment (note the right-most vertex):*

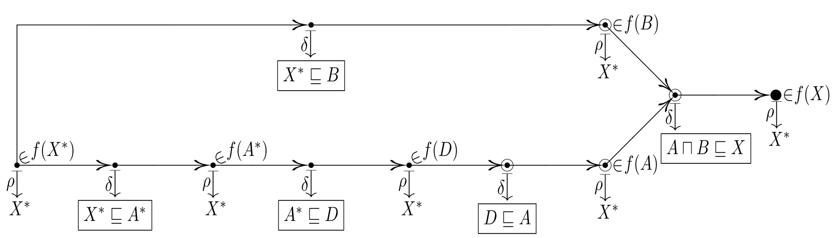

Figure 4. Constructing concept interpolant from a subgraph in enriched model

Consider the vertex labelled with $f(A)$ and X^* in the graph, denote it as v. It corresponds to the entailed inclusion $X^* \sqsubseteq A$ and is special, because it contains non-Δ-symbols from $\mathcal{T}^*$ and $\mathcal{T}$ on the LHS and RHS, respectively, while all the descending vertices contain non-Δ-symbols only from $\mathcal{T}^*$. By the concept interpolation property of $\mathcal{EL}$, there must be an interpolant Δ-concept for the *inclusion $X^* \sqsubseteq A$ (for the vertex v). In fact, this concept can be obtained from the LHS of the axiom labeling the child vertex of v, i.e. D is the required interpolant for v. To obtain the interpolant for the right-most vertex (corresponding to the inclusion $X^* \sqsubseteq X$ of interest) it suffices to take the conjunction of interpolants for the child vertices preceding the one labeled with $A \sqcap B \sqsubseteq X$. We have $B \in \Delta$ and the concept $D \sqcap B$ is the required interpolant.*

The example shows that there are special intermediate derived concept inclusions in the proof for which an interpolant can be easily computed. It is either the LHS of the side condition of the rule under which the concept inclusion is obtained, or the concept from the RHS of one of the premises of the rule. In particular, such special concept inclusions are the very first formulas in branches of the proof which are derived using axioms of both, a TBox $\mathcal{T}$ and its "copy". For the rest of the derived concept inclusions, the corresponding interpolants can be obtained by combining interpolants for preceding concept inclusions under conjunction and existential restriction. These are the main ideas presented in the proof of Theorem 2 and the procedure of interpolant computation. Motivated by the observation on these "special" concept inclusions in proofs, we first introduce the following auxiliary notion.

Definition 5 (Irregular formula, irregular vertex). *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be normal TBoxes, with $\mathtt{sig}\,(\mathcal{T}_1) \cap \mathtt{sig}\,(\mathcal{T}_2) = \Delta$ for a signature Δ. A formula φ is called regular (wrt $\mathcal{T}_1, \mathcal{T}_2$) if $\mathtt{sig}\,(\varphi) \subseteq \mathtt{sig}\,(\mathcal{T}_i)$ for some $i = 1, 2$. Otherwise, it is called irregular. Let $((\Gamma^+, \Gamma^-, E), \delta, \rho, f)$ be an enriched model of $\mathcal{T}_1 \cup \mathcal{T}_2$. A vertex $v^+ \in \Gamma^+$ is called irregular if*

- *either $\rho(v) = A$, $v \in f(B)$, for concept names A, B, and $A \sqsubseteq B$ is an irregular formula,*
- *or $\rho(v) = \langle A, B\rangle$, $v \in f(r)$, for concept names A, B, and $A \sqsubseteq \exists r.B$ is an irregular formula.*

Theorem 2 (Structural proof of concept interpolation). *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be normal TBoxes, with $\mathtt{sig}\,(\mathcal{T}_1) \cap \mathtt{sig}\,(\mathcal{T}_2) = \Delta$ for a signature Δ, and let $A \sqsubseteq F$ be a normal, but irregular formula such that $A \in \mathtt{sig}\,(\mathcal{T}_1)$. If there is an inference of $A \sqsubseteq F$ from $\mathcal{T}_1 \cup \mathcal{T}_2$, then there is a concept interpolant D such that $\mathcal{T}_1 \cup \mathcal{T}_2 \models \{A \sqsubseteq D,\ D \sqsubseteq F\}$ and $\mathtt{sig}\,(D) \subseteq \Delta$.*

Proof. We apply induction on the length of the inference of $A \sqsubseteq F$ from $\mathcal{T}_1 \cup \mathcal{T}_2$. Observe that by the form of the rules in Figure 1, we must have $F \in \mathtt{sub}\,(\mathcal{T}_1 \cup \mathcal{T}_2)$ and thus, $\mathtt{sig}\,(F) \subseteq \mathtt{sig}\,(\mathcal{T}_2)$, since $A \sqsubseteq F$ is irregular and $A \in \mathtt{sig}\,(\mathcal{T}_1)$. Induction step (the base case is trivial): let R be the corresponding derivation rule. Note that the formula in the side condition of R is regular and assume that all the formulas in the premise are regular too. If R is $Trans$, then we conclude that B is the required interpolant; if R is $AndR$, then $A_1 \sqcap ... \sqcap A_n$ is; if R is Ex, then either $\{A, r, A_1, A_2\} \subseteq \mathtt{sig}\,(\mathcal{T}_1)$ and thus, $\exists r.A_2$ is the required interpolant, or $\{A_1, r, A_2, B\} \subseteq \mathtt{sig}\,(\mathcal{T}_2)$, and hence, $\exists r.A_1$ is.

Now assume there is an irregular formula in the premise of R. By the induction hypothesis, there is a corresponding interpolant for this formula. If R is $Trans$, then it is the required interpolant for $A \sqsubseteq F$. If R is $AndR$, then consider the maximal subset $I \subseteq \{1,..,n\}$ such that for every $i \in I$ the formula $A \sqsubseteq A_i$ is irregular and D_i is the corresponding interpolant. Since $A \sqsubseteq F$ is irregular, we conclude that $A_j \in \Delta$ for all $j \in \{1,..,n\} \setminus I$ and therefore, $\bigsqcap_{i\in I} D_i \sqcap \bigsqcap_{j\in\{1,..,n\}\setminus I} A_j$ is the required interpolant. Finally, if R is Ex then, in case $A \sqsubseteq \exists r.A_1$ is irregular, its interpolant is the required interpolant for $A \sqsubseteq F$. Now assume that $A \sqsubseteq \exists r.A_1$ is a regular formula and $A_1 \sqsubseteq A_2$ is irregular with an interpolant D. Then $r \in \Delta$ and $\exists r.D$ is the required interpolant for $A \sqsubseteq F$. □

The proof demonstrates that in case an irregular formula φ is derived from $\mathcal{T}_1 \cup \mathcal{T}_2$ by an inference rule R, then either there are only regular formulas in the premise of R and the interpolant for φ can be computed immediately, or there are some irregular formulas in the premise for which the corresponding interpolants must be computed first in order to obtain an interpolant for φ. Practically, this means a recursive procedure of computing interpolants while traversing the proof of formula φ (i.e. a submodel starting with the plus-vertex corresponding to φ). Every proof starts from axioms which are regular formulas, so the first irregular formula in a branch of the proof is obtained after application of some rule R with regular premises. As noted in the proof of Theorem 2 (and Example 4), an interpolant in this case can be computed immediately, since it is either the concept from the LHS of the formula in the side condition of R, or the concept $\exists r.A_1$ from the RHS of the premise, if $\mathsf{R} = Ex$. Thus, Theorem 2 (together with Theorem 1) evidences soundness of the recursive procedure given below, while termination follows from the finiteness of the graph in an enriched model and the usage of flags for visited vertices. The procedure is defined in words instead of pseudo-code for ease of exposition.

Procedure for computing an explicit definition as concept interpolant
Input:

- an enriched model $\mathcal{M}^* = (\Gamma = (\Gamma^+, \Gamma^-, E), \delta, \rho, f)$ of the union of normal TBoxes $\mathcal{T}_1 \cup \mathcal{T}_2$, with $\mathtt{sig}(\mathcal{T}_1) \cap \mathtt{sig}(\mathcal{T}_2) = \Delta$, for a signature Δ;
- a concept inclusion $A_1 \sqsubseteq A_2$, with $A_i \in (\mathtt{sig}(\mathcal{T}_i) \cap \mathsf{N_c}) \setminus \Delta$, for $i = 1, 2$.

Output: Empty iff $\mathcal{T}_1 \cup \mathcal{T}_2 \not\models A_1 \sqsubseteq A_2$; otherwise, a concept interpolant D such that $\mathtt{sig}(D) \subseteq \Delta$ and $\mathcal{T}_1 \cup \mathcal{T}_2 \models \{A_1 \sqsubseteq D,\ D \sqsubseteq A_2\}$.

The procedure operates on the graph Γ from $\mathcal{M}^*$ and incorporates a subroutine of graph traversal.

1. If there is no vertex $v^+ \in \Gamma^+$ such that $v^+ \in f(A_2)$ and $\rho(v^+) = A_1$, then the procedure terminates with the empty output. Otherwise, initiate graph traversal starting from the vertex v^+ as follows.

2. Initialize two maps $vis : \Gamma^+ \cup \Gamma^- \to \{0, 1\}$ and $int : \Gamma^+ \cup \Gamma^- \to Term(\Delta) \cup \{null\}$ for all vertices v of Γ as: $vis(v) = 0$, $int(v) = null$. The map vis will be used to mark visited vertices and int will mark vertices with their corresponding interpolants. The value *null* will mean that there is no interpolant and $vis(v) = 0$ will imply $int(v) = null$. The task is compute the value of $int(v^+)$, which will not be equal to *null*, however it can be the case for other vertices in Γ (which are related to proofs inside TBoxes $\mathcal{T}_1$, $\mathcal{T}_2$, but not from their union, i.e. to the proofs consisting of regular formulas).

3. The recursive subrootine of graph traversal and interpolant computation is as follows. Let v be the currently visited vertex from Γ.

3.1. Let $v \in \Gamma^+$. Then we visit an arbitrary vertex w such that $(w, v) \in E$, i.e. a child minus-vertex:

- If $vis(w) = 1$, we set $int(v) = int(w)$;
- If $vis(w) = 0$, then we initiate recursively graph traversal starting from the vertex w. When the traversal is complete, we set $int(v) = int(w)$.

After these steps, we set $vis(v) = 1$, i.e. visiting the vertex v is finished and we exit the recursion to one level up.

3.2. Let $v \in \Gamma^-$. Then we immediately set $vis(v) = 1$. Let $\delta(v) = F_1 \sqsubseteq F_2$. If for some $i = 1, 2$, we have $\texttt{sig}(F_i) \subseteq \Delta$, then we set $int(v) = F_i$ and exit the recursion to one level up.

Otherwise, initiate graph traversals recursively starting from every irregular (plus-vertex) w, with $(w, v) \in E$. If there are no such vertices, then exit the recursion to one level up. Otherwise, when the traversal is complete for all such vertices w, we have $int(w) \neq null$ and proceed as follows.

3.3. Consider different cases for the formula $\delta(v)$.

- If $\delta(v) = A \sqsubseteq B$, or $\delta(v) = A \sqsubseteq \exists r.B$, for some concept names A, B, then there is a single child plus-vertex w for v and we set $int(v) = int(w)$.
- If $\delta(v) = A_1 \sqcap ... \sqcap A_n \sqsubseteq B$, $n > 1$, then let $w_1, .., w_n$ be the child plus-vertices for v, with $w_i \in f(A_i)$ and $A = \rho(w_i)$, for $i = 1, ..n$ and a concept name A. Take the maximal subset $I \subseteq \{1, .., n\}$ such that for every $i \in I$, the vertex w_i is irregular and set $int(v) = \bigsqcap_{i \in I} int(w_i) \ \sqcap \ \bigsqcap_{j \in \{1,..,n\} \setminus I} A_j$.
- If $\delta(v) = \exists r.A_2 \sqsubseteq B$, then there is a child plus-vertex w such that $w \in f(r)$ and $\rho(v) = \langle A, A_1 \rangle$ for some A and A_1. If w is irregular, then set $int(v) = int(w)$. Otherwise there is an irregular child vertex w, with $w \in f(A_2)$, $\rho(w) = A_1$, and we set $int(v) = \exists r.int(w)$.

This finishes visiting the vertex $v \in \Gamma^-$.

The definition of the procedure is complete.

Let v be the minus-vertex from point 3.3 of the procedure and w be one of its child plus-vertices with an interpolant $int(w) = D$ for some concept D. From the consideration of the formula $\delta(v)$, one can see that either the concept $int(v)$ is the same as $int(w)$, or can be of length increased by a constant (when $\delta(v) = \exists r.A_2 \sqsubseteq B$), or in the worst case, $int(v)$ can be n-times longer than $int(w)$ (if $\delta(v) = A_1 \sqcap ... \sqcap A_n \sqsubseteq B$, $n > 1$, and each corresponding child plus-vertex w_i is irregular). This means that the size of the interpolant for the inclusion $A_1 \sqsubseteq A_2$ is at worst $\mathcal{O}(n^m)$, where m is the number of axioms in the normal TBox $\mathcal{T}_1 \cup \mathcal{T}_2$ and n is the maximal number of conjuncts in an axiom of $\mathcal{T}_1 \cup \mathcal{T}_2$. Note that due to arbitrary selection of the minus-vertex in point 3.1, the procedure computes an arbitrary chosen interpolant out of existing ones for the concept inclusion $A_1 \sqsubseteq A_2$. By Theorem 1, it should be clear that point 3.1 can be easily modified to decide concept definability wrt a subset S of the TBox: it suffices to consider only those child minus-vertices which are labeled with formulas corresponding to axioms of S.

5 Partial Concept Reformulation with Enriched Models

The problem of partial reformulation of concepts under a TBox $\mathcal{T}$ wrt a signature Δ considered in the introduction implies several instances of the concept definability problem for different extensions of Δ. At worst, the number of tests for definability to be performed is equal to the number of signature symbols in a concept C of interest. Moreover, in the decomposition algorithm for $\mathcal{EL}$ [6], each non-primitive concept being the LHS of an axiom of $\mathcal{T}$ has to be tested for partial reformulation. Each instance of the definability problem wrt an extension $\Delta' \supseteq \Delta$ requires checking entailment from the union of TBoxes $(\mathcal{T} \setminus S) \cup (\mathcal{T}' \setminus S')$, where $\mathcal{T}'$ is a "copy" of $\mathcal{T}$, with all non-Δ'-symbols injectively renamed into fresh symbols, and S, S' are subsets of axioms of $\mathcal{T}$ and $\mathcal{T}'$, respectively. Since the concept C and the sets $(\Delta' \setminus \Delta)$, S, S' are the changing parameters in the problem of partial reformulation, but the TBox $\mathcal{T}$ and the initial signature Δ are fixed, it makes sense to use a prebuilt extended representation of $\mathcal{T}$ to be able to solve numerous instances of the concept definability problem efficiently. In the following we argue that an enriched model precomputed for a given TBox $\mathcal{T}$ and a signature Δ can be used as a template for solving different instances of the concept definability problem wrt the changing parameters.

We assume w.l.o.g. that the input TBox $\mathcal{T}$ is in the enriched normal form. Given $\mathcal{T}$ and Δ, first an enriched model $\mathcal{M}$ for $\mathcal{T}$ is built and then its "copy" $\mathcal{M}^*$ is generated wrt an injective renaming of all non-Δ-symbols from $\mathcal{T}$. The pair $\mathcal{M}, \mathcal{M}^*$ is then kept as a precomputed template. As soon as an extension Δ' of the signature Δ is given, the graphs of these two models are expanded (with the iterative process of model construction from Section 3) wrt a set of formulas generated from $\mathcal{T}$ and the set $\Delta' \setminus \Delta$. The idea of constructing this set of formulas is described in Example 5 and Theorem 3.

In the example below, we assume that we are given TBoxes $\mathcal{T}$ and $\mathcal{T}^*$ sharing a signature Δ (initially taken to be empty for simplicity) and then Δ is expanded as the result of substituting (*or "gluing"*) some symbols of $\mathcal{T}^*$ with that of $\mathcal{T}$.

Example 5 *Consider normal TBoxes* $\mathcal{T} = \{A \sqsubseteq \exists r_1.A_1\}$ *and* $\mathcal{T}^* = \{\exists s_1.B_1 \sqsubseteq B\}$*, we have* $\Delta = \mathtt{sig}(\mathcal{T}) \cap \mathtt{sig}(\mathcal{T}^*) = \varnothing$*. Let* $\mathcal{T}' = \{\exists r_1.A_1 \sqsubseteq B\}$ *be the TBox obtained from* $\mathcal{T}^*$ *by substituting* B_1 *with* A_1 *and* s_1 *with* r_1*; we have* $\Delta' = \mathtt{sig}(\mathcal{T}) \cap \mathtt{sig}(\mathcal{T}') = \{A_1, r_1\} \supseteq \Delta$*. Clearly,* $\mathcal{T} \cup \mathcal{T}' \models A \sqsubseteq B$*.*

Now consider the TBox $\mathcal{Q} = \mathcal{T} \cup \mathcal{T}^* \cup \{A_1 \equiv B_1\} \cup \mathcal{T}^*(s/r)$*, where* $\mathcal{T}^*(s/r) = \{\exists r_1.B_1 \sqsubseteq B\}$*, i.e. this TBox consists of the axiom of* $\mathcal{T}^*$ *with the role* s_1 *substituted with* r_1*. Then* $\mathcal{Q} \models A \sqsubseteq B$ *and in fact, it will be shown in Theorem 3 that* $\mathcal{T} \cup \mathcal{T}'$ *and the theory* $\mathcal{Q}$ *constructed for the union* $\mathcal{T} \cup \mathcal{T}^*$ *in the way above entail the same normal formulas in signature* $\mathtt{sig}(\mathcal{T} \cup \mathcal{T}^*)$*. Therefore, an enriched model for checking entailment from the union of theories* $\mathcal{T} \cup \mathcal{T}'$ *(sharing the signature* $\Delta' \supseteq \Delta$*) can be obtained by expanding prebuilt models for* $\mathcal{T}$ *and* $\mathcal{T}^*$ *wrt the additional formulas* $\{A_1 \equiv B_1\} \cup \mathcal{T}^*(s/r)$*.*

We note that a much more general model-theoretic result implying Theorem 3 can be proved, but here we restrict ourselves only to the case important in the context of this paper.

Theorem 3 (Formulas imitate substitutions of signature symbols). *Let* $\mathcal{T}_1$ *and* $\mathcal{T}_2$ *be normal TBoxes,* $\Delta = \mathtt{sig}(\mathcal{T}_1) \cap \mathtt{sig}(\mathcal{T}_2)$*, and let* $\sigma_1 = \{A_i\}_{i \in I} \cup$

$\{r_j\}_{j\in J} \subseteq \mathtt{sig}(\mathcal{T}_1) \setminus \Delta$, $\sigma_2 = \{B_i\}_{i\in I} \cup \{s_j\}_{j\in J} \subseteq \mathtt{sig}(\mathcal{T}_2) \setminus \Delta$ *be signatures, where all* $A_i, B_i \in \mathsf{N_c}$ *and* $r_j, s_j \in \mathsf{N_r}$. *Let* $\mathcal{T}_2'$ *be the result of substituting every* B_i *with* A_i *and every* s_j *with* r_j *in the axioms of* $\mathcal{T}_2$.

Consider the TBox $\mathcal{Q} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \{A_i \equiv B_i\}_{i\in I} \cup \mathcal{T}_2(\bar{s}/\bar{r})$, *where* $\mathcal{T}_2(\bar{s}/\bar{r})$ *is the set of formulas of the form* $\exists r_j.X \sqsubseteq Y$, *or* $X \sqsubseteq \exists r_j.Y$, $j \in J$ *such that* $\exists s_j.X \sqsubseteq Y$, *or* $X \sqsubseteq \exists s_j.Y$, *respectively, is an axiom of* $\mathcal{T}_2$. *Then for any normal formula* φ *in signature* $\mathtt{sig}(\mathcal{T}_1 \cup \mathcal{T}_2')$ *we have* $\mathcal{T}_1 \cup \mathcal{T}_2' \models \varphi$ *iff* $\mathcal{Q} \models \varphi$.

Proof. ($\Rightarrow$): We show that $\mathcal{Q}$ is a conservative extension of $\mathcal{T}_1 \cup \mathcal{T}_2'$. By the definition of $\mathcal{Q}$ it suffices to verify that $\mathcal{Q} \models \mathcal{T}_2' \setminus \mathcal{T}_2$. Let $\psi \in \mathcal{T}_2' \setminus \mathcal{T}_2$, this is a formula obtained from an axiom $\xi \in \mathcal{T}_2$ by substituting B_i with A_i and s_j with r_j for $i \in I$, $j \in J$. Let $\{A_k, \ldots, A_l\} = \mathtt{sig}(\psi) \cap \sigma_1 \cap \mathsf{N_c}$ (a possibly empty set). Consider several possible cases for the signature of ψ. If $\mathtt{sig}(\psi) \cap \{r_j\}_{j\in J} = \varnothing$, then $\xi \cup \{A_i \equiv B_i\}_{i\in\{k,\ldots,l\}} \models \psi$, so we have $\mathcal{Q} \models \psi$, because $\xi \in \mathcal{T}_2 \subseteq \mathcal{Q}$. If there is a role $r_j \in \mathtt{sig}(\psi)$ for some $j \in J$, then there must be a formula $\gamma \in \mathcal{T}_2(\bar{s}/\bar{r})$ such that either $\gamma \models \psi$ (in case $\mathtt{sig}(\psi) \cap \sigma_1 \cap \mathsf{N_c} = \varnothing$), or $\{\gamma\} \cup \{A_i \equiv B_i\}_{i\in\{k,\ldots,l\}} \models \psi$. In each of the cases we have $\mathcal{Q} \models \psi$.

($\Leftarrow$): Let $\mathcal{Q}'$ be the TBox obtained from $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_2(\bar{s}/\bar{r})$ by substituting every occurrence of symbol B_i with A_i, for $i \in I$. We have $\mathcal{Q} \models \varphi$ iff $\mathcal{Q}' \models \varphi$, since $\mathcal{Q} \models \mathcal{Q}'$, every model of $\mathcal{Q}'$ expands to a model of $\mathcal{Q}$, and $\mathtt{sig}(\varphi) \subseteq \mathtt{sig}(\mathcal{Q}')$. Observe that any formula $\exists r_j.X \sqsubseteq Y$, or $X \sqsubseteq \exists r_j.Y$ from a "copy" of $\mathcal{T}_2(\bar{s}/\bar{r})$ obtained after this substitution, is a formula of $\mathcal{T}_2'$. Therefore, $\mathcal{T}_1 \cup \mathcal{T}_2' \subseteq \mathcal{Q}'$ and the set difference $\mathcal{D} = \mathcal{Q}' \setminus (\mathcal{T}_1 \cup \mathcal{T}_2')$ consists of formulas of the form $\exists s_j.X \sqsubseteq Y$, or $X \sqsubseteq \exists s_j.Y$, with $X, Y \in \mathtt{sig}(\mathcal{T}_2')$, $j \in J$, such that there exists a corresponding axiom $\exists r_j.X \sqsubseteq Y$, or $X \sqsubseteq \exists r_j.Y$ in $\mathcal{T}_2'$. We now prove by induction that for an inference of a (normal) formula φ from $\mathcal{Q}'$, with $\mathtt{sig}(\varphi) \subseteq \mathtt{sig}(\mathcal{T}_1 \cup \mathcal{T}_2')$, there is a corresponding inference of φ from $\mathcal{T}_1 \cup \mathcal{T}_2'$. Induction step (the base case is trivial): if φ is obtained by *AndR* or *Trans*, then by the form of formulas in $\mathcal{D}$, the axiom in the side condition can not belong to $\mathcal{D}$ and thus there is nothing to prove. If the rule is *Ex* with an axiom $\exists s_j.Y' \sqsubseteq Z \in \mathcal{D}$ in the side condition, then there are formulas of the form $X \sqsubseteq \exists s_j.Y \in \mathcal{D}$ and $Y \sqsubseteq Y'$ in the premise. The formula $Y \sqsubseteq Y'$ is normal and $Y, Y' \in \mathtt{sig}(\mathcal{T}_2')$, thus by the induction hypothesis, there exists an inference of $Y \sqsubseteq Y'$ from $\mathcal{T}_1 \cup \mathcal{T}_2'$. Since there are corresponding formulas $\exists r_j.Y' \sqsubseteq Z$ and $X \sqsubseteq \exists r_j.Y$ in $\mathcal{T}_2'$, we conclude that there is an inference of φ from $\mathcal{T}_1 \cup \mathcal{T}_2'$ in which φ is obtained by *Ex* from formulas $X \sqsubseteq \exists r_j.Y$, $Y \sqsubseteq Y'$ and $X \sqsubseteq \exists r_j.Y$ is obtained by *Trans* from the tautology $X \sqsubseteq X$. $\square$

It should be clear that precomputing the models $\mathcal{M}, \mathcal{M}^*$ does make sense, when there is a large number of queries for partial concept reformulation to be executed and answering *all* these queries would potentially involve all the axioms of the TBox. In paticular, this is the case for the decomposition algorithm from [6]. On the other hand, an expansion of $\mathcal{M}, \mathcal{M}^*$ wrt the auxiliary axioms from Theorem 3 may be computed only to the extent needed for checking the entailment $C' \sqsubseteq C$ for a concept C of interest (and its copy C' after a renaming wrt $\Delta' \supseteq \Delta$).

6 Discussion

In the paper, we did not consider questions of efficient implementation of enriched models and the algorithm of their construction. The iterative procedure based on the rules from Section 3 (including the initialization step) is mentioned to simplify the exposition and is by no means proposed to be used in applications; this topic is left out of the scope of this paper to separate the main concepts from their implementation details. We note that the technique of enriched models can be easily adopted for solving the problem of concept reformulation by using the inference systems from the consequence-driven procedures for $\mathcal{EL}$ (see e.g. [4] or [5]). Moreover, we argue that the basic ideas described in this paper can be implemented in the context of any reasoning system for $\mathcal{EL}$ which allows for proof tracing and incremental reasoning (see a short overview in Section 2 of [5]). The first feature is necessary for constructing concept interpolants from proofs, while the second one allows for using a prebuilt model of a TBox and computing its local changes after the TBox is updated. For assume we are given a TBox $\mathcal{T}$, a signature $\Delta \subseteq \mathcal{T}$, and we would like to be able to solve different instances of the partial concept reformulation problem under $\mathcal{T}$ wrt Δ. Let $\mathcal{T}^*$ be a copy of $\mathcal{T}$, with *all* non-Δ-symbols injectively renamed, and let $\mathfrak{M}$ be a prebuilt model for $\mathcal{T} \cup \mathcal{T}^*$. Let A^* and $\mathcal{S}^*$ be the corresponding copies of A and $\mathcal{S}$ in $\mathcal{T}^*$. Now assume a signature $\Delta' \supseteq \Delta$ and a set $\mathcal{S} \subseteq \mathcal{T}$ are given; take here $\Delta' \setminus \Delta = \{A\}$, $A \in \mathsf{N_c}$, for simplicity. Then building a model for the TBox $(\mathcal{T} \setminus \mathcal{S}) \cup (\mathcal{T}' \setminus \mathcal{S}')$ (where $\mathcal{T}'$ and $\mathcal{S}'$ are copies of $\mathcal{T}$ and $\mathcal{S}$ under an injective renaming of all non-Δ'-symbols) is equivalent to incrementally updating $\mathfrak{M}$ after removing the axioms of $\mathcal{S} \cup \mathcal{S}^*$ from $\mathcal{T} \cup \mathcal{T}^*$ and adding the axiom $A \equiv A^*$.

Since an enriched model of a TBox contains proofs of primitive concept inclusions, it can be directly used for computing concept interpolants and hence, explicit definitions of concepts. For every proof of a concept inclusion $C^* \sqsubseteq C$ from $\mathcal{T} \cup \mathcal{T}^*$ (where C^* and $\mathcal{T}^*$ are appropriate copies of C and $\mathcal{T}$), there is a corresponding explicit definition of C. It follows from the example on the non-succinctness of uniform interpolants from Section 7 of [7] that, given a $\mathcal{EL}$-TBox $\mathcal{T}$ and a primitive concept inclusion φ, there can be exponentially many proofs of φ from $\mathcal{T}$ and this holds already for the case, when $\mathtt{sig}(\mathcal{T}) \subseteq \mathsf{N_c}$. Consider the signature $\Delta = \{B_{ij} \mid B_{ij} \in \mathsf{N_c},\ 1 \leqslant i,j \leqslant n\}$ and the TBox $\mathcal{T} = \{A \sqsubseteq B_{ij} \mid 1 \leqslant i,j \leqslant n\} \cup \{B_{ij} \sqsubseteq B_i \mid 1 \leqslant i,j \leqslant n\} \cup \{B_1 \sqcap ... \sqcap B_n \sqsubseteq A\}$. Then the concept A has the longest explicit Δ-definition under $\mathcal{T}$ which is $\bigsqcap_{1 \leqslant i,j \leqslant n} B_{ij}$, as well as exponentially many shortest Δ-definitions. The procedure of computing concept interpolants given in Section 4 implies traversing only a single proof out of the available ones and the choice of the proof is not regulated by any strategies. We believe that the development of strategies of concept reformulation deserves attention not only in the Description Logic $\mathcal{EL}$, but also in more expressive languages.

Acknowledgments. The first author was supported by the German Research Foundation within the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems", the Russian Academy of Sciences (Grant No. 15/10), and the Siberian Division of the Russian Academy of Sciences (Integration Project No. 3). The second author was supported by the Russian Foundation for Basic Research (Grant No. 12-01-00460-a).

References

1. Baader F., Brand S., and Lutz C.: Pushing the EL Envelope. In Proc. Nineteenth International Joint Conference on Artificial Intelligence (IJCAI 2005).
2. ten Cate B., Conradie W., Marx M., Venema Y.: Definitorially Complete Description Logics. In Proc. Tenth International Conference on Principles of Knowledge Representation and Reasoning (KR 2006).
3. ten Cate B., Franconi E., Seylan I.: Beth Definability in Expressive Description Logics. In Proc. Twenty-second International Joint Conference on Artificial Intelligence (IJCAI 2011).
4. Kazakov Y., Krötzsch M., and Simancik F. Concurrent Classification of EL Ontologies. In: Proc. Tenth International Semantic Web Conference (ISWC 2011).
5. Kazakov Y. and Klinov P. Incremental Reasoning in OWL EL without Bookkeeping. Technical report, University of Ulm, May 2013. Available from: http://code.google.com/p/elk-reasoner/
6. Konev B., Lutz C., Ponomaryov D., Wolter F.: Decomposing Description Logic Ontologies. In: Proc. Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010).
7. Konev B., Walther D., and Wolter F. The Logical Difference Problem for Description Logic Terminologies. In: Proc. Fourth International Joint Conference on Automated Reasoning (IJCAR 2008).
8. Seylan I., Franconi E., de Bruijn J.: Effective Query Rewriting with Ontologies over Dboxes. In: Proc. Twenty-first International Joint Conference on Artificial Intelligence (IJCAI 2009).