Computing Minimal Projection Modules for Conjunctive Queries*

Jieying Chen[1], Michel Ludwig[2], Yue Ma[1], and Dirk Walther[3]

[1] Laboratoire de Recherche en Informatique, Université Paris-Sud, France
{jieying.chen,yue.ma}@lri.fr
[2] Luxembourg, michel.ludwig@gmail.com
[3] Fraunhofer IVI, Dresden, Germany
dirk.walther@ivi.fraunhofer.de

Abstract. We consider the problem of extracting modules of an ontology that contains the knowledge as represented by a second ontology. The knowledge to be preserved is specified using entailment of conjunctive queries over a given vocabulary. We propose a novel module notion called *projection module* that preserves the answers to conjunctive queries as they follow from a reference ontology. We present an algorithm for computing minimal projection modules for conjunctive queries. As target and reference ontology we take $\mathcal{ELH}^r$-terminologies. The algorithm is based on simulation notions developed for detecting logical differences between $\mathcal{ELH}^r$-terminologies.

1 Introduction

Ontology comparison can help understanding the overlap and differences among ontologies which is often desired while a user manipulates multiple knowledge sources. In this paper, we propose the notion of *projection module* which allows to compare the entailment capacities of two ontologies about a given vocabulary. A projection module characterizes the relative knowledge of one ontology by taking another one as a reference. This can thus lead to a fine-grained ontology comparison measurement between two ontologies.

Various approaches to comparing ontologies have been suggested, including ontology mapping or alignment [12], and logical difference [19–21, 23]. Ontology matching is the process of determining correspondences, in particular, the subsumption, equivalence, or disjointness relations between two concept or role names from different ontologies. A good concept similarity [1, 22] is often helpful for ontology matching. In contrast, logical difference focuses on the comparison of entailed logical consequences from each ontology and returns difference witnesses if differences are present.

When an ontology has no logical difference compared to another one, our approach further extracts sub-ontologies of the first ontology that contain the

* This work is partially funded by the ANR project GoAsQ (ANR-15-CE23-0022).

knowledge as represented by the second ontology. For example, suppose $\mathcal{T}_1 = \{A_1 \sqsubseteq A_2, A_2 \sqsubseteq A_3\}$, $\mathcal{T}_2 = \{A_1 \sqsubseteq A_3 \sqcap B_1, B_1 \sqsubseteq \exists r.A_3\}$, and $\Sigma = \{A_1, A_3, r\}$. Then $\mathcal{T}_2$ entails all queries about Σ that follow from $\mathcal{T}_1$. However, the projection module of $\mathcal{T}_2$ with respect to $\mathcal{T}_1$ and Σ consists of $A_1 \sqsubseteq A_3 \sqcap B_1$. This means that a strict sub-ontology of $\mathcal{T}_2$ is sufficient to capture all the information of $\mathcal{T}_1$ about Σ. Moreover, $\mathcal{T}_2$ entails the consequence $A_1 \sqsubseteq \exists r.A_3$, which is not the case for $\mathcal{T}_1$. Intuitively, $\mathcal{T}_2$ is richer in information about Σ than $\mathcal{T}_1$.

Ontology modularity [9, 16, 19, 21, 24, 25] is about the extraction of sub-ontologies that preserve all logical consequences over a given signature. In contrast, the proposed projection module is different from modules of a single ontology. For the example above, the basic minimal module [9], $\bot\top^\star$-module, and MEX-module [18] of T_2 w.r.t. Σ are all $\mathcal{T}_2$ itself. The extra axiom in $\mathcal{T}_2$ compared to its projection module with respect to $\mathcal{T}_1$ confirms that $\mathcal{T}_1$ conveys less information about Σ than $\mathcal{T}_2$, which gives a way to compare these two ontologies. For latest results on logical inseparability (in particular, inseparability w.r.t. conjunctive queries), see [7,8,13], and for a survey on query inseparability, see [6].

To compute projection modules, in this paper, we generalize the notion of justification to the notion of *subsumption justification* as a minimal set of axioms that maintains a consequence. Our algorithm employs the classical notion of justification to compute subsumption justification. Currently, the approaches for computing all the justifications of an ontology w.r.t. a consequence can be classified into two categories: "glass-box" [2,5,14,15] and "black-box" [10,14,26].

We proceed as follows. Section 2 gives a brief review of Description Logic $\mathcal{EL}$ and its extensions as well as logical difference. In Section 3, we introduce the notion of project module. In Section 4, the notion and the computation of role subsumption justifications are presented, which are employed in Section 5 to compute project modules. Finally, Section 6 concludes the paper.

2 Preliminaries

We start by briefly reviewing the description logic $\mathcal{EL}$ and several of its extension with range restrictions and conjunction of roles and the universal role as well as concept subsumptions based on these extensions. For a more detailed introduction to description logics, we refer to [3, 4].

Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be sets of concept names and role names. We assume these sets to be mutually disjoint and countably infinite. The sets of *$\mathcal{EL}$-concepts* C, *$\mathcal{EL}^{\mathsf{ran}}$-concepts* D, *$\mathcal{EL}^{\sqcap}$-concepts* E, and $\mathcal{EL}^{\sqcap,u}$-concepts F, and the sets of *$\mathcal{ELH}^r$-inclusions* α and *$\mathcal{EL}^{\mathsf{ran},\sqcap,u}$-inclusions* β are built according to the grammar rules:

$$
\begin{aligned}
C &::= A \mid C \sqcap C \mid \exists r.C \mid \mathsf{dom}(r)\\
D &::= A \mid D \sqcap D \mid \exists r.D \mid \mathsf{dom}(r) \mid \mathsf{ran}(r)\\
E &::= A \mid E \sqcap E \mid \exists R.E\\
F &::= A \mid F \sqcap F \mid \exists R.F \mid \exists u.F\\
\alpha &::= C \sqsubseteq C \mid \mathsf{ran}(r) \sqsubseteq C \mid \mathsf{ran}(r) \sqcap C \sqsubseteq C \mid C \equiv C \mid r \sqsubseteq s\\
\beta &::= D \sqsubseteq F \mid r \sqsubseteq s
\end{aligned}
$$

where $A \in \mathsf{N_C}$, $r, s \in \mathsf{N_R}$, u is a fresh logical symbol (the *universal role*) and $R = r_1 \sqcap \ldots \sqcap r_n$ with $r_1, ..., r_n \in \mathsf{N_R}$, for $n \geq 1$. We refer to inclusions also as *axioms*. A Γ*-TBox* is a finite set of Γ-inclusions, where Γ ranges over the sets of $\mathcal{ELH}^r$- and $\mathcal{EL}^{\mathsf{ran},\sqcap,u}$-inclusions.

The semantics is defined as usual in terms of interpretations, which interpret concept and role names and are inductively extended to complex concepts. The notions of satisfaction of a concept, axiom and TBox as well as the notions of a model and the logical consequence relation are defined as usual. We skip a detailed introduction here.

A signature is a finite set of symbols from $\mathsf{N_C}$ and $\mathsf{N_R}$. We write $\mathrm{sig}^{\mathsf{N_C}}(\xi)$ and $\mathrm{sig}(\xi)$ for the set of concept names and the set of concept and role names occurring in a syntactic object ξ. The symbol Σ is used as a subscript to a set of concepts or inclusions to denote that the elements only use symbols from Σ.

An $\mathcal{ELH}^r$*-terminology* $\mathcal{T}$ is an $\mathcal{ELH}^r$-TBox consisting of axioms α of the form $A \sqsubseteq C$, $A \equiv C$, $r \sqsubseteq s$, $\mathsf{ran}(r) \sqsubseteq C$ or $\mathsf{dom}(r) \sqsubseteq C$, where A is a concept name, C an $\mathcal{EL}$-concept and no concept name occurs more than once on the left-hand side of an axiom.[1] To simplify the presentation we assume that terminologies do not contain axioms of the form $A \equiv B$ or $A \equiv \top$ (after removing multiple $\top$-conjuncts) for concept names A and B. For a terminology $\mathcal{T}$, let $\prec_{\mathcal{T}}$ be a binary relation over $\mathsf{N_C}$ such that $A \prec_{\mathcal{T}} B$ iff there is an axiom of the form $A \sqsubseteq C$ or $A \equiv C$ in $\mathcal{T}$ such that $B \in \mathrm{sig}(C)$. A terminology $\mathcal{T}$ is *acyclic* if the transitive closure $\prec_{\mathcal{T}}^{+}$ of $\prec_{\mathcal{T}}$ is irreflexive; otherwise $\mathcal{T}$ is *cyclic*. A concept name A is said to be *conjunctive in* $\mathcal{T}$ iff there exist concept names $B_1, \ldots, B_n$, $n > 0$, such that $A \equiv B_1 \sqcap \ldots \sqcap B_n \in \mathcal{T}$; otherwise A is said to be *non-conjunctive in* $\mathcal{T}$.

An $\mathcal{ELH}^r$-terminology $\mathcal{T}$ is *normalised* iff it only contains axioms of the forms $\varphi \sqsubseteq B_1 \sqcap \ldots \sqcap B_n$, $A \sqsubseteq \exists r.B$, $A \sqsubseteq \mathsf{dom}(r)$, $r \sqsubseteq s$, and $A \equiv B_1 \sqcap \ldots \sqcap B_m$, $A \equiv \exists r.B$, where $\varphi \in \{A, \mathsf{dom}(s), \mathsf{ran}(s)\}$, $n \geq 1$, $m \geq 2$, $A, B, B_i \in \mathsf{N_C}$, $r, s \in \mathsf{N_R}$, and each conjunct B_i is non-conjunctive in $\mathcal{T}$. Every $\mathcal{ELH}^r$-terminology $\mathcal{T}$ can be normalised in polynomial time such that the resulting terminology is a conservative extension of $\mathcal{T}$ [17]. A subset $M \subseteq \mathcal{T}$ is called a *justification for an* $\mathcal{ELH}$*-concept inclusion* α *from* $\mathcal{T}$ iff $M \models \alpha$ and $M' \not\models \alpha$ for every $M' \subsetneq M$. We denote the set of all justifications for an $\mathcal{ELH}$-concept inclusion α from an $\mathcal{ELH}$-terminology $\mathcal{T}$ with $\mathrm{Just}_{\mathcal{T}}(\alpha)$. The latter may contain exponentially many justifications in the number of axioms in $\mathcal{T}$.

We now recall the notion of logical difference for concept subsumption queries, instance queries and conjunctive queries from [17, 20].

Definition 1 (Logical Difference). *The* $\mathcal{EL}^{\mathsf{ran},\sqcap,u}$*-subsumption query and conjunctive query difference between* $\mathcal{T}_1$ *and* $\mathcal{T}_2$ *wrt.* Σ *are the sets* $\mathsf{cDiff}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$ *and* $\mathsf{qDiff}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$*, where*

- $\varphi \in \mathsf{cDiff}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$ *iff* φ *is an* $\mathcal{EL}^{\mathsf{ran},\sqcap,u}_{\Sigma}$*-inclusion and* $\mathcal{T}_1 \models \varphi$ *and* $\mathcal{T}_2 \not\models \varphi$*;*
- $(\mathcal{A}, q(\boldsymbol{a})) \in \mathsf{qDiff}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$ *iff* $\mathcal{A}$ *is a* Σ*-ABox and* $q(\boldsymbol{a})$ *a* Σ*-conjunctive query such that* $(\mathcal{T}_1, \mathcal{A}) \models q(\boldsymbol{a})$ *and* $(\mathcal{T}_2, \mathcal{A}) \not\models q(\boldsymbol{a})$.

[1] A concept equation $A \equiv C$ stands for the inclusions $A \sqsubseteq C$ and $C \sqsubseteq A$.

The following theorem states that $\mathcal{EL}^{\mathsf{ran},\sqcap,u}$-subsumption queries are sufficient to detect the absence of conj. query differences (Lemmas 62 and 63 in [17]).

Theorem 1. $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$ *iff* $\mathsf{qDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$.

If the set $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ is not empty, then it typically contains infinitely many concept inclusion. We make use of the *primitive witnesses theorems* from [17], which state that if there is a concept inclusion difference in $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$, then there exists an inclusion in $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ of one of the following three types $\delta_1, \delta_2, \delta_3$, which are built according to the grammar rules below:

$$\begin{aligned}
\delta_1 &::= r \sqsubseteq s \\
\delta_2 &::= D \sqsubseteq A \\
\delta_3 &::= A \sqsubseteq E \mid \mathsf{dom}(r) \sqsubseteq E \mid \mathsf{ran}(r) \sqsubseteq E
\end{aligned}$$

where δ_1 ranges over role inclusions, δ_2 is an $\mathcal{EL}^{\mathsf{ran}}$-inclusion, and δ_3 is an $\mathcal{EL}^{\mathsf{ran},\sqcap,u}$-inclusion. Note that each of these inclusions has either a simple left-hand or a simple right-hand side.

The set of all $\mathcal{EL}^{\mathsf{ran},\sqcap,u}$-subsumption difference witnesses is defined as

$$\mathsf{Wtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2) := (\mathsf{roleWtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2), \mathsf{lhsWtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2), \mathsf{rhsWtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)),$$

where the set $\mathsf{roleWtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ consists of all type-δ_1 inclusions in $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$, and the sets $\mathsf{lhsWtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2) \subseteq (\Sigma \cap \mathsf{N_C}) \cup \{ \mathsf{dom}(r) \mid r \in \Sigma \} \cup \{ \mathsf{ran}(r) \mid r \in \Sigma \}$ and $\mathsf{rhsWtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2) \subseteq \mathsf{N_C} \cap \Sigma$ of *left-hand* and *right-hand subsumption query difference witnesses* consist of the left-hand sides of the type-δ_3 inclusions in $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ and the right-hand sides of type-δ_2 inclusions in $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$, respectively. Consequently, the set $\mathsf{Wtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ can be seen as a finite representation of the set $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ [17], which is typically infinite. As a corollary of the primitive witness theorems in [17], we have that the representation is complete in the following sense: $\mathsf{cDiff}_\Sigma(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$ iff $\mathsf{Wtn}_\Sigma(\mathcal{T}_1, \mathcal{T}_2) = (\emptyset, \emptyset, \emptyset)$. Thus, deciding the existence of concept inclusion differences is equivalent to deciding non-emptiness of the three witness sets.

3 Projection Modules

A terminology $\mathcal{T}_1$ together with a signature Σ and a query language $\mathcal{Q}$ determine a set Φ of queries from $\mathcal{Q}$ formulated using only symbols from Σ that follow from $\mathcal{T}_1$. Intuitively, Φ captures the $\mathcal{Q}$-knowledge of $\mathcal{T}_1$ about Σ. In this paper, we consider conjunctive queries or, equivalently, $\mathcal{EL}^{\mathsf{ran},\sqcap,u}$-subsumption queries as a query language. A projection module for Φ of another terminology $\mathcal{T}_2$ is a subset of $\mathcal{T}_2$ that entails the queries in Φ. The $\mathcal{Q}$-knowledge in Φ of $\mathcal{T}_1$ about Σ is captured by the projection module of $\mathcal{T}_2$ and it is represented using axioms from $\mathcal{T}_2$. The projection can be seen as transforming the finite representation of Φ with axioms from $\mathcal{T}_1$ into a finite representation of Φ with axioms from $\mathcal{T}_2$. As an application, the projection of Φ onto $\mathcal{T}_2$ allows us to determine the axioms that implement Φ in $\mathcal{T}_2$. For instance, projection modules enable the verification of compliance to certain modelling guidelines and standards by the axioms of $\mathcal{T}_2$ that implement Φ. In particular, projection modules that are minimal w.r.t. set inclusion are relevant for this task. Note that existing module notions suggest to compute a module of $\mathcal{T}_2$ for the given signature, Σ, to obtain the $\mathcal{Q}$-knowledge of $\mathcal{T}_2$ about Σ, whereas we are only interested in Φ, i.e. the $\mathcal{Q}$-knowledge of $\mathcal{T}_1$ about Σ. Consequently, extracting modules of $\mathcal{T}_2$ for Σ will yield modules that generally contain irrelevant axioms and that are, therefore, likely too large for manual inspection.

Definition 2 (Projection Module). *Let* $\rho = \langle \mathcal{T}_1, \Sigma, \mathcal{T}_2 \rangle$ *be a* projection setting. *A set* $\mathcal{M} \subseteq \mathcal{T}_2$ *is a* conjunctive query projection module under ρ *iff for every* Σ*-ABox* $\mathcal{A}$ *and every* Σ*-conjunctive query* $q(\boldsymbol{a})$*:* $(\mathcal{T}_1, \mathcal{A}) \models q(\boldsymbol{a})$ *implies* $(\mathcal{M}, \mathcal{A}) \models q(\boldsymbol{a})$.

There may exist several (even exponentially many) minimal projection modules.

Example 1. Let $\mathcal{T}_1 = \{A_1 \sqsubseteq A_4\}$, $\mathcal{T}_2 = \{A_1 \equiv A_2 \sqcap A_3, A_2 \sqsubseteq A_4, A_3 \sqsubseteq A_4\}$ and $\Sigma = \{A_1, A_4\}$. Then the conjunctive query projection module under $\rho = \langle \mathcal{T}_1, \Sigma, \mathcal{T}_2 \rangle$ is $\mathcal{T}_2$.

The notion of projection module is not symmetric, i.e., a projection module under $\langle \mathcal{T}_1, \Sigma, \mathcal{T}_2 \rangle$ is not necessarily the same as a projection module under $\langle \mathcal{T}_2, \Sigma, \mathcal{T}_1 \rangle$. When the reference terminology equals the terminology from which axioms are to be extracted, a reflexive projection setting of the form $\rho^{\circlearrowleft} = \langle \mathcal{T}, \Sigma, \mathcal{T} \rangle$ is used. We call a projection module under $\rho^{\circlearrowleft}$ also an *automorphic projection module*.

An interesting application of the projection module is to compare entailment capacities of two terminologies, as shown in the following example.

Example 2. Let $\mathcal{T}_1 = \{A_i \sqsubseteq A_{i+1} \mid 1 \leq i \leq n\}$, $\mathcal{T}_2 = \{A_1 \sqsubseteq A_n, B_1 \sqsubseteq B_2\}, \Sigma = \{A_1, A_n, B_1, B_2\}$. Intuitively, $\mathcal{T}_1$ contains less information about Σ than $\mathcal{T}_2$. The minimal projection module under $\rho = \langle \mathcal{T}_1, \Sigma, \mathcal{T}_2 \rangle$ is $\mathcal{M} = \{A_1 \sqsubseteq A_n\}$. By using $|\mathcal{M}|/|\mathcal{T}_2| = 1/2$ as a measure, we see that only half of $\mathcal{T}_2$ is about the information of $\mathcal{T}_1$ w.r.t. Σ.

4 Representing Projection Modules using Justifications

In this section, we introduce justification notions for sets of inclusions and we show how they can be combined to obtain minimal projection modules. We start with defining the notion of role subsumption justifications for a set of $\mathcal{EL}^{\mathsf{ran},\sqcap,u}$-inclusions of the form $r \sqsubseteq s$, where $r, s \in \Sigma$ (cf. Section 2).

Definition 3 (Role Subsumption Justification). *Let* $\rho = \langle \mathcal{T}_1, \Sigma, \mathcal{T}_2 \rangle$. *A set* $\mathcal{M}$ *is called a* role subsumption module under ρ *iff* $\mathcal{M} \subseteq \mathcal{T}_2$ *and for every* $r, s \in \mathsf{N_R} \cap \Sigma$, $\mathcal{T}_1 \models r \sqsubseteq s$ *implies* $\mathcal{M} \models r \sqsubseteq s$. *A* role subsumption justification under ρ *is the role subsumption module under* ρ *that is minimal w.r.t.* $\subsetneq$.

We denote the set of all role subsumption justifications under ρ as $\mathcal{J}_\rho^R$.

Lemma 1. *Let* $J \in \mathcal{J}_\rho^R$. *Then* $\mathsf{roleWtn}_\Sigma(\mathcal{T}_1, J) = \emptyset$.

We continue with defining the notion of subsumption justifications for inclusions of $\mathcal{EL}^{\mathsf{ran},\sqcap,u}$ that are of the form $D \sqsubseteq F$, where D ranges over $\mathcal{EL}^{\mathsf{ran}}$-concepts and F over $\mathcal{EL}^{\sqcap,u}$-concepts.

Definition 4 (Subsumption Justification). *A subsumption setting is a tuple* $\chi = \langle \mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2 \rangle$, *where* $\mathcal{T}_1$ *and* $\mathcal{T}_2$ *are normalised* $\mathcal{ELH}^r$*-terminologies,* Σ *is a signature,* $X_1, X_2 \in \mathsf{N_C} \cup \{\, \mathsf{dom}(r), \mathsf{ran}(r) \mid r \in \mathsf{N_R} \,\}$.

A set $\mathcal{M}$ *is called a* subsumer module under χ *iff* $\mathcal{M} \subseteq \mathcal{T}_2$ *and for every* $C \in \mathcal{EL}_\Sigma^{\mathsf{ran},\sqcap,u}$, $\mathcal{T}_1 \models X_1 \sqsubseteq C$ *implies* $\mathcal{M} \models X_2 \sqsubseteq C$. $\mathcal{M}$ *is called a* subsumee module under χ *iff* $\mathcal{M} \subseteq \mathcal{T}_2$ *and for every* $C \in \mathcal{EL}_\Sigma^{\mathsf{ran},\sqcap,u}$, $\mathcal{T}_1 \models C \sqsubseteq X_1$ *implies* $\mathcal{M} \models C \sqsubseteq X_2$.

$\mathcal{M}$ *is called a* subsumption module under χ *iff* $\mathcal{M}$ *is a subsumer module, a subsumee module under* χ *and role subsumption module under* $\langle \mathcal{T}_1, \Sigma, \mathcal{T}_2 \rangle$. *A* subsumee (resp. subsumer, subsumption) justification under χ *is a subsumee (resp. subsumer, subsumption) module under* χ *that is minimal w.r.t.* $\subsetneq$.

We denote the set of all subsumee (resp. subsumer, subsumption) justifications under χ as $\mathcal{J}_\chi^{\leftarrow}$ (resp. $\mathcal{J}_\chi^{\rightarrow}$, $\mathcal{J}_\chi$), where $\chi = \langle \mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2 \rangle$.

Using Definition 1 and 4, we obtain the following proposition stating the absence of certain concept names, and domain and range restrictions of role names as left-hand and right-hand difference witnesses between a reference terminology $\mathcal{T}_1$ and a subsumer and subsumee justification of a second terminology $\mathcal{T}_2$.

For a signature Σ, let $\Sigma^{\mathsf{dom}} = \{\, \mathsf{dom}(t) \mid t \in \mathsf{N_R} \cap \Sigma \,\}$ and $\Sigma^{\mathsf{ran}} = \{\, \mathsf{ran}(t) \mid t \in \mathsf{N_R} \cap \Sigma \,\}$ be the sets consisting of concepts of the form $\mathsf{dom}(t)$ and $\mathsf{ran}(t)$ for every role name t in Σ, respectively. Furthermore, let $\Sigma^\zeta = \Sigma \cup \Sigma^{\mathsf{dom}} \cup \Sigma^{\mathsf{ran}}$ for $\zeta \in \mathcal{C}^\Sigma$.

Lemma 2. *Let* $\varphi \in (\Sigma \cap \mathsf{N_C}) \cup \Sigma^{\mathsf{dom}} \cup \Sigma^{\mathsf{ran}}$ *and let* $A \in \Sigma \cap \mathsf{N_C}$. *Additionally, let* $\chi = \langle \mathcal{T}_1, \varphi, \Sigma, \mathcal{T}_2, \varphi \rangle$ *and* $\chi' = \langle \mathcal{T}_1, A, \Sigma, \mathcal{T}_2, A \rangle$. *Then:*

- $\varphi \notin \mathsf{lhsWtn}_\Sigma(\mathcal{T}_1, J_\chi)$ *for every* $J_\chi \in \mathcal{J}_\chi^{\rightarrow}$;
- $A \notin \mathsf{rhsWtn}_\Sigma(\mathcal{T}_1, J_{\chi'})$ *for every* $J_{\chi'} \in \mathcal{J}_{\chi'}^{\leftarrow}$.

To obtain subsumption modules we can use an operator $\otimes$ to combine sets of role, subsumer and subsumee justifications, one justification for each potential difference witness that needs to be prevented; cf. Lemmas 1 and 2. Given a set S and $\mathbb{S}_1, \mathbb{S}_2 \subseteq 2^S$, $\mathbb{S}_1 \otimes \mathbb{S}_2 := \{\, S_1 \cup S_2 \mid S_1 \in \mathbb{S}_1, S_2 \in \mathbb{S}_2 \,\}$. For instance, if $\mathbb{S}_1 = \{\{\alpha_1, \alpha_2\}, \{\alpha_3\}\}$ and $\mathbb{S}_2 = \{\{\alpha_1, \alpha_3\}, \{\alpha_4, \alpha_5\}\}$, then $\mathbb{S}_1 \otimes \mathbb{S}_2 = \{\{\alpha_1, \alpha_2, \alpha_3\}, \{\alpha_1, \alpha_2, \alpha_4, \alpha_5\}, \{\alpha_3, \alpha_4, \alpha_5\}, \{\alpha_1, \alpha_3\}\}$. For a set $\mathbb{M}$ of sets, we define a function $\mathrm{Minimise}_\subseteq(\mathbb{M})$ as follows: $\mathcal{M} \in \mathrm{Minimise}_\subseteq(\mathbb{M})$ iff $\mathcal{M} \in \mathbb{M}$ and there does not exist a set $\mathcal{M}' \in \mathbb{M}$ such that $\mathcal{M}' \subsetneq \mathcal{M}$. Continuing the previous example, $\mathrm{Minimise}_\subseteq(\mathbb{S}_1 \otimes \mathbb{S}_2) = \{\{\alpha_2, \alpha_3\}, \{\alpha_1, \alpha_2, \alpha_4, \alpha_5\}, \{\alpha_3, \alpha_4, \alpha_5\}\}$.

We now use $\otimes$ and $\mathrm{Minimise}_\subseteq(\cdot)$ to combine sets of role, subsumer and subsumee justifications to obtain the set of all minimal projection modules.

Theorem 2. *Let $\mathbb{M}_\rho$ be the set of all projection modules under $\rho = \langle \mathcal{T}_1, \Sigma, \mathcal{T}_2 \rangle$ that are minimal w.r.t. $\subsetneq$. Then the following holds, where $\chi(\psi) = \langle \mathcal{T}_1, \psi, \Sigma, \mathcal{T}_2, \psi \rangle$:*

$$\mathbb{M}_\rho = Minimize_{\subseteq}\big(\mathcal{J}^R_\rho \otimes \bigotimes_{\varphi \in (\Sigma \cap \mathsf{N_C}) \cup \Sigma^{\mathsf{dom}} \cup \Sigma^{\mathsf{ran}}} \mathcal{J}^{\rightarrow}_{\chi(\varphi)} \otimes \bigotimes_{A \in \Sigma \cap \mathsf{N_C}} \mathcal{J}^{\leftarrow}_{\chi(A)} \big)$$

5 Computing Projection Modules

In this section, we present algorithms for computing role, subsumer and subsumee justifications. The algorithms use the following notion of a cover of a set of sets. For a finite set S and a set $\mathbb{T} \subseteq 2^S$, we say that a set $\mathbb{M} \subseteq 2^S$ is a *cover of* $\mathbb{T}$ iff $\mathbb{M} \subseteq \mathbb{T}$ and for every $\mathcal{M} \in \mathbb{T}$, there exists $\mathcal{M}' \in \mathbb{M}$ such that $\mathcal{M}' \subseteq \mathcal{M}$. In other words, a cover is a subset of $\mathbb{T}$ containing all sets from $\mathbb{T}$ that are minimal w.r.t. $\subsetneq$. Therefore, a cover of the set of all subsumption modules also contains all subsumption justifications. We will use covers to characterise the output of our algorithms to ensure that all justifications have been computed.

Algorithm 7 shows how to collect relevant Σ-role inclusions for role subsumption justifications (cf. Def. 3). The following proposition states its correctness.

Proposition 1. $\mathcal{J}^R_\rho = \textsc{Cover}^R(\mathcal{T}_1, \Sigma, \mathcal{T}_2)$*, where* $\rho = \langle \mathcal{T}_1, \Sigma, \mathcal{T}_2 \rangle$.

5.1 Computing Subsumer Justifications

We now present our algorithm for computing subsumer justifications. The algorithm relies on the notion of a subsumer simulation between terminologies from [11, 23]. For defining the simulation notion, we need a specific notion of reachability. For $\varphi \in \mathsf{N_C} \cup \{ \mathsf{dom}(r), \mathsf{ran}(r) \mid r \in \mathsf{N_R} \}$ and a normalised $\mathcal{ELH}^r$-terminology $\mathcal{T}$, let $F_{\mathcal{T}}(\varphi)$ be the smallest set closed under the following three conditions: $\varphi \in F_{\mathcal{T}}(\varphi)$; $Y \in F_{\mathcal{T}}(\varphi)$ if $X \in F_{\mathcal{T}}(\varphi)$, $\mathcal{T} \models X \sqsubseteq X'$ and $X' \bowtie \exists r.Y \in \mathcal{T}$; and $\mathsf{dom}(r) \in F_{\mathcal{T}}(\varphi)$ if $\mathsf{ran}(r) \in F_{\mathcal{T}}(\varphi)$.

Definition 5 (Subsumer Simulation). *A relation $S \subseteq N(\mathcal{T}_1, \Sigma) \times N(\mathcal{T}_2, \Sigma)$, where $N(\mathcal{T}, \Sigma) = \{ X, \mathsf{dom}(r), \mathsf{ran}(r) \mid X, r \in (sig(\mathcal{T}) \cup \Sigma), X \in \mathsf{N_C}, r \in \mathsf{N_R} \}$, is a* Σ-subsumer simulation from $\mathcal{T}_1$ to $\mathcal{T}_2$ *iff the following conditions are satisfied:*

- ($S^{\rightarrow}_{\mathsf{N_C}}$) *if $(X_1, X_2) \in S$, then for every $\varphi \in \Sigma \cup \Sigma^{\mathsf{dom}}$ with $\mathcal{T}_1 \models X_1 \sqsubseteq \varphi$, it holds that $\mathcal{T}_2 \models X_2 \sqsubseteq \varphi$;*
- ($S^{\rightarrow}_{\exists}$) *if $(X_1, X_2) \in S$ and $X'_1 \bowtie_1 \exists r.Y_1 \in \mathcal{T}_1$ with $\bowtie_1 \in \{\sqsubseteq, \equiv\}$ such that $\mathcal{T}_1 \models X_1 \sqsubseteq X'_1$ and $\mathcal{T}_1 \models r \sqsubseteq s$ for some $s \in \Sigma$, there exists $X'_2 \bowtie_2 \exists r'.Y_2 \in \mathcal{T}_2$ with $\bowtie_2 \in \{\sqsubseteq, \equiv\}$ such that $\mathcal{T}_2 \models X_2 \sqsubseteq X'_2$ and for every $s \in \Sigma$ with $\mathcal{T}_1 \models r \sqsubseteq s$, it holds that $\mathcal{T}_2 \models r' \sqsubseteq s$ and $(Y_1, Y_2) \in S$.*

We write $\mathcal{T}_1 \sim^{\rightarrow}_{\Sigma} \mathcal{T}_2$ iff there exists a Σ-subsumer simulation S from $\mathcal{T}_1$ to $\mathcal{T}_2$ such that for every $\varphi \in (\Sigma \cap \mathsf{N_C}) \cup \Sigma^{\mathsf{dom}} \cup \Sigma^{\mathsf{ran}}$: $(\varphi, \varphi) \in S$, and for every $\psi_1 \in F_{\mathcal{T}_1}(\varphi)$, there exists a $\psi_2 \in F_{\mathcal{T}_2}(\varphi)$ such that $(\psi_1, \psi_2) \in S$.

For $X_1, X_2 \in \mathsf{N_C}$, we write $\langle \mathcal{T}_1, X_1 \rangle \sim^{\rightarrow}_{\Sigma} \langle \mathcal{T}_2, X_2 \rangle$ iff there exists a Σ-subsumer simulation S from $\mathcal{T}_1$ to $\mathcal{T}_2$ with $(X_1, X_2) \in S$ for which $\mathcal{T}_1 \sim^{\rightarrow}_{\Sigma} \mathcal{T}_2$.

A subsumer simulation conveniently captures the set of subsumers in the following sense: If a Σ-subsumer simulation from $\mathcal{T}_1$ to $\mathcal{T}_2$ contains the pair (X_1, X_2), then X_2 entails w.r.t. $\mathcal{T}_2$ all subsumers of X_1 w.r.t. $\mathcal{T}_1$ that are formulated in the signature Σ. Formally, we obtain the following theorems from [23].

Theorem 3. *It holds that* $\mathcal{T}_1 \sim_{\Sigma}^{\rightarrow} \mathcal{T}_2$ *iff* $\mathsf{lhsWtn}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$.

Theorem 4. *Let* $\langle \mathcal{T}_1, X_1 \rangle \sim_{\Sigma}^{\rightarrow} \langle \mathcal{T}_2, X_2 \rangle$. *Then for every* $D \in \mathcal{EL}_{\Sigma}^{\mathsf{ran},\sqcap,u}$: $\mathcal{T}_1 \models X_1 \sqsubseteq D$ *implies* $\mathcal{T}_2 \models X_2 \sqsubseteq D$.

Algorithm 2 collects all the axioms necessary to satisfy the conditions of Definition 5. Observe that $\textsc{Cover}_{\rightarrow}(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2)$ may be called several times during the execution of the algorithm. A possible optimisation is to store return values in memory in order to retrieve them more quickly for subsequent calls.

The following theorem shows that Algorithm 2 indeed computes the set of subsumer modules, thus producing a cover of subsumer justifications.

Theorem 5. *Let* $\chi = \langle \mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2 \rangle$ *and* $\mathbb{M} := \textsc{Cover}_{\rightarrow}^{cq}(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2)$. *If* $\mathcal{T}_1 \sim_{\Sigma}^{\rightarrow} \mathcal{T}_2$, *then* $\mathbb{M}$ *is a cover of the set of subsumer justifications under* χ.

5.2 Computing Subsumee Justifications

Next we present the algorithm for computing subsumee justifications based on a notion of a subsumee simulation. The basic idea of the algorithm is to collect as few axioms from $\mathcal{T}_2$ as possible to maintain the subsumee simulation between Σ-concept names.

First we present some auxiliary notions for handling conjunctions on the left-hand side of subsumptions. We define for each concept name X a so-called *definitorial forest* consisting of sets of axioms of the form $Y \equiv Y_1 \sqcap \ldots \sqcap Y_n$ which can be thought of as forming *trees*. Any subsumee justification under $\langle \mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2 \rangle$ contains the axioms of a selection of these trees, i.e., one tree for every conjunction formulated over Σ that entails X_1 w.r.t. $\mathcal{T}_1$. Formally, we define a set of a $\mathrm{DefForest}_{\mathcal{T}}^{\sqcap}(X) \subseteq 2^{\mathcal{T}}$ to be the smallest set closed under the following conditions: $\emptyset \in \mathrm{DefForest}_{\mathcal{T}}^{\sqcap}(X)$; $\{\alpha\} \in \mathrm{DefForest}_{\mathcal{T}}^{\sqcap}(X)$ for $\alpha = X \equiv X_1 \sqcap \ldots \sqcap X_n \in \mathcal{T}$; and $\Gamma \cup \{\alpha\} \in \mathrm{DefForest}_{\mathcal{T}}^{\sqcap}(X)$ for $\Gamma \in \mathrm{DefForest}_{\mathcal{T}}^{\sqcap}(X)$ with $Z \equiv Z_1 \sqcap \ldots \sqcap Z_k \in \Gamma$ and $\alpha = Z_i \equiv Z_i^1 \sqcap \ldots \sqcap Z_i^n \in \mathcal{T}$. Given $\Gamma \in \mathrm{DefForest}_{\mathcal{T}}^{\sqcap}(X)$, we set $\mathrm{leaves}(\Gamma) := \mathrm{sig}(\Gamma) \setminus \{\, X \in \mathrm{sig}(C) \mid X \equiv C \in \Gamma \,\}$ if $\Gamma \neq \emptyset$; and $\{X\}$ otherwise. We denote the maximal element of $\mathrm{DefForest}_{\mathcal{T}}^{\sqcap}(X)$ w.r.t. $\subseteq$ with $\text{max-tree}_{\mathcal{T}}^{\sqcap}(X)$. Finally, we set $\text{non-conj}_{\mathcal{T}}(X) := \mathrm{leaves}(\text{max-tree}_{\mathcal{T}}^{\sqcap}(X))$.

For example, let $\mathcal{T} = \{\alpha_1, \alpha_2, \alpha_3\}$, where $\alpha_1 := X \equiv Y \sqcap Z$, $\alpha_2 := Y \equiv Y_1 \sqcap Y_2$, and $\alpha_3 := Z \equiv Z_1 \sqcap Z_2$. Then $\mathrm{DefForest}_{\mathcal{T}}^{\sqcap}(X) = \{\emptyset, \{\alpha_1\}, \{\alpha_1, \alpha_2\}, \{\alpha_1, \alpha_3\}, \{\alpha_1, \alpha_2, \alpha_3\}\}$. We have that $\mathrm{leaves}(\{\alpha_1, \alpha_3\}) = \{Y, Z_1, Z_2\}$, $\text{max-tree}_{\mathcal{T}}^{\sqcap}(X) = \{\alpha_1, \alpha_2, \alpha_3\}$, and $\text{non-conj}_{\mathcal{T}}(X) = \{Y_1, Y_2, Z_1, Z_2\}$.

We say that a concept name A is Σ-entailed w.r.t. $\mathcal{T}$ iff there is an $\mathcal{EL}_{\Sigma}^{\mathsf{ran}}$-concept C such that $\mathcal{T} \models C \sqsubseteq A$; and we say that a role name s is Σ-entailed in $\mathcal{T}$ iff there exists $s' \in \mathsf{N_R} \cap \Sigma$ such that $\mathcal{T} \models s' \sqsubseteq s$.

We now define the notion of a *subsumee simulation* from $\mathcal{T}_1$ to $\mathcal{T}_2$ as a subset of $\text{sig}^{\mathsf{N_C}}(\mathcal{T}_1) \times \text{sig}^{\mathsf{N_C}}(\mathcal{T}_2) \times \mathcal{C}^{\Sigma}_{\mathcal{T}_1}$, where $\mathcal{C}^{\Sigma}_{\mathcal{T}_1} \coloneqq \{\epsilon\} \cup (\mathsf{N_R} \cap (\Sigma \cup \text{sig}(\mathcal{T}_1)))$ is the range of role contexts.

Definition 6 (Subsumee Simulation). *A relation $S \subseteq sig^{\mathsf{N_C}}(\mathcal{T}_1) \times sig^{\mathsf{N_C}}(\mathcal{T}_2) \times \mathcal{C}^{\Sigma}_{\mathcal{T}_1}$ is a Σ-*subsumee simulation from $\mathcal{T}_1$ to $\mathcal{T}_2$ *iff the following conditions hold:*

($S^{\leftarrow}_{\mathsf{N_C}}$) *if $(X_1, X_2, \zeta) \in S$, then for every $\varphi \in \Sigma^{\zeta}$ and for every $X_2' \in$ non-conj$_{\mathcal{T}_2}(X_2)$ with $\mathcal{T}_2 \not\models \mathsf{ran}(\zeta) \sqsubseteq X_2'$, $\mathcal{T}_1 \models \varphi \sqsubseteq X_1$ implies $\mathcal{T}_2 \models \varphi \sqsubseteq X_2'$;*

($S^{\leftarrow}_{\exists}$) *if $(X_1, X_2, \zeta) \in S$ and $X_1 \equiv \exists r.Y_1 \in \mathcal{T}_1$ such that $\mathcal{T}_1 \models s \sqsubseteq r$ for $s \in \Sigma$ and Y_1 is Σ-entailed w.r.t. $\mathcal{T}_1$, then for every $X_2' \in$ non-conj$_{\mathcal{T}_2}(X_2)$ not entailed by $\mathsf{dom}(s)$ or $\mathsf{ran}(\zeta)$ w.r.t. $\mathcal{T}_2$, there exists $X_2' \equiv \exists r'.Y_2 \in \mathcal{T}_2$ such that $\mathcal{T}_2 \models s \sqsubseteq r'$ and $(Y_1, Y_2, s) \in S$;*

($S^{\leftarrow}_{\sqcap}$) *if $(X_1, X_2, \zeta) \in S$ and $X_1 \equiv Y_1 \sqcap \ldots \sqcap Y_n \in \mathcal{T}_1$, then for every $Y_2 \in$ non-conj$_{\mathcal{T}_2}(X_2)$ not entailed by $\mathsf{ran}(\zeta)$ in $\mathcal{T}_2$, there exists $Y_1 \in$ non-conj$_{\mathcal{T}_1}(X_1)$ not entailed by $\mathsf{ran}(\zeta)$ w.r.t. $\mathcal{T}_2$ such that $(Y_1, Y_2, \epsilon) \in S$.*

We write $\mathcal{T}_1 \sim^{\leftarrow}_{\Sigma} \mathcal{T}_2$ iff there is a Σ-subsumer simulation S from $\mathcal{T}_1$ to $\mathcal{T}_2$ such that for every $A \in \Sigma \cap \mathsf{N_C}$: $(A, A, \epsilon) \in S$.

For $\zeta \in \Sigma \cap \mathsf{N_R}$, we write $\langle \mathcal{T}_1, X_1 \rangle \sim^{\leftarrow}_{\Sigma,\zeta} \langle \mathcal{T}_2, X_2 \rangle$ iff there is a Σ-subsumer simulation S from $\mathcal{T}_1$ to $\mathcal{T}_2$ with $(X_1, X_2, \zeta) \in S$ for which $\mathcal{T}_1 \sim^{\leftarrow}_{\Sigma} \mathcal{T}_2$.

Analogously to subsumer simulations, a subsumee simulation captures the set of subsumees as it is made precise in the following theorems.

Theorem 6. $\mathcal{T}_1 \sim^{\leftarrow}_{\Sigma} \mathcal{T}_2$ *iff* $\mathsf{rhsWtn}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$.

Theorem 7. *Let $\langle \mathcal{T}_1, X_1 \rangle \sim^{\leftarrow}_{\Sigma,\zeta} \langle \mathcal{T}_2, X_2 \rangle$. Then for every $C \in \mathcal{EL}^{\mathsf{ran}}_{\Sigma}$: $\mathcal{T}_1 \models C \sqsubseteq X_1$ implies that $\mathcal{T}_2 \models C \sqsubseteq X_2$.*

Before introducing the algorithms, we first extend the notion of Σ-entailment. We say that a concept name X is *complex Σ-entailed w.r.t. $\mathcal{T}$* iff for every $Y \in$ non-conj$_{\mathcal{T}}(X)$ one of the following conditions holds:

- there exists $B \in \Sigma$ such that $\mathcal{T} \models B \sqsubseteq Y$ and $\mathcal{T} \not\models B \sqsubseteq X$; or
- there exists $Y \equiv \exists r.Z \in \mathcal{T}$ and r, Z are Σ-entailed in $\mathcal{T}$.

Otherwise, X is said to be *simply Σ-entailed*. For example, let $\mathcal{T} = \{X \equiv X_1 \sqcap X_2,\ B_1 \sqsubseteq X_1,\ X_2 \equiv \exists r.Z,\ B_2 \sqsubseteq Z, s \sqsubseteq r\}$. We have that non-conj$_{\mathcal{T}}(X) = \{X_1, X_2\}$, then r is Σ-entailed w.r.t. $\mathcal{T}$; X is complex Σ-entailed w.r.t. $\mathcal{T}$ for $\Sigma = \{B_1, B_2, s\}$; but X is not complex Σ'-entailed w.r.t. $\mathcal{T}$, where Σ' ranges over $\{B_1, B_2\}$, $\{B_1, s\}$, $\{B_2, s\}$. Additionally, X is not complex Σ-entailed w.r.t. $\mathcal{T} \cup \{B_1 \sqsubseteq X\}$.

Algorithm 3 is responsible for computing a cover of all subsumee justifications. It orchestrates three further algorithms in order to collect the axioms necessary to satisfy the three conditions of Definition 6: Algorithm 4 for Case ($S^{\leftarrow}_{\mathsf{N_C}}$), Algorithm 5 for Case ($S^{\leftarrow}_{\exists}$) and Algorithm 6 for Case ($S^{\leftarrow}_{\sqcap}$). While Algorithm 4 can readily be understood, we provide some additional explanation for the remaining algorithms.

Algorithm 1: Computing a Cover of all Subsumer Justifications for Conjunctive Queries

function $\textsc{Cover}^{\mathsf{cq}}_{\rightarrow}$ $(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2)$
  $\mathbb{M}^{\rightarrow}_{(X_1,X_2)} \coloneqq \{\emptyset\}$
  for every $\psi_1 \in F_{\mathcal{T}_1}(X_1)$ **do**
    $\mathbb{M}^{\rightarrow}_{\psi_1} \coloneqq \{\emptyset\}$
    for every $\psi_2 \in F_{\mathcal{T}_2}(X_2)$ *such that* $\langle \mathcal{T}_1, \psi_1\rangle \sim^{\rightarrow}_{\Sigma} \langle \mathcal{T}_2, \psi_2\rangle$ **do**
      $\mathbb{M}^{\rightarrow}_{\psi_1,\psi_2} \coloneqq \textsc{Cover}_{\rightarrow}(\mathcal{T}_1, \psi_1, \Sigma, \mathcal{T}_2, \psi_2)$
      $\mathbb{M}^{\rightarrow}_{\psi_1} \coloneqq \mathbb{M}^{\rightarrow}_{\psi_1} \cup \mathbb{M}^{\rightarrow}_{\psi_1,\psi_2}$
    $\mathbb{M}^{\rightarrow}_{(X_1,X_2)} = \mathbb{M}^{\rightarrow}_{(X_1,X_2)} \otimes \mathbb{M}^{\rightarrow}_{\psi_1}$
  return $\text{Minimise}_{\subseteq}(\mathbb{M}^{\rightarrow}_{(X_1,X_2)})$

Algorithm 2: Computing a Cover of all Subsumer Justifications (Recursive)

function $\textsc{Cover}_{\rightarrow}$ $(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2)$
  $\mathbb{M}^{\rightarrow}_{(X_1,X_2)} \coloneqq \{\emptyset\}$
  for every $B \in (\Sigma \cap \mathsf{N_C}) \cup \{\, \mathsf{dom}(r) \mid r \in \Sigma \,\}$ *such that* $\mathcal{T}_1 \models X_1 \sqsubseteq B$ **do**
    $\mathbb{M}^{\rightarrow}_{(X_1,X_2)} \coloneqq \mathbb{M}^{\rightarrow}_{(X_1,X_2)} \otimes \text{Just}_{\mathcal{T}_2}(X_2 \sqsubseteq B)$
  for every $Y \bowtie_1 \exists r.Z \in \mathcal{T}_1$ $(\bowtie_1 \in \{\sqsubseteq, \equiv\})$ *with* $\mathcal{T}_1 \models X_1 \sqsubseteq Y$, $\mathcal{T}_1 \models r \sqsubseteq s$ *for some* $s \in \Sigma \cap \mathsf{N_R}$ **do**
    $\mathbb{M}^{\rightarrow}_{\exists r.Z} \coloneqq \{\emptyset\}$
    for every $Y' \bowtie_2 \exists r'.Z' \in \mathcal{T}_2$ $(\bowtie_2 \in \{\sqsubseteq, \equiv\})$ *with* $\mathcal{T}_2 \models X_2 \sqsubseteq Y'$ *and* $\mathcal{T}_2 \models r' \sqsubseteq s$ *for every* $s \in \{\, s' \in \Sigma \cap \mathsf{N_R} \mid \mathcal{T}_1 \models r \sqsubseteq s' \,\}$ *and* $\langle \mathcal{T}_1, Z\rangle \sim^{\rightarrow}_{\Sigma} \langle \mathcal{T}_2, Z'\rangle$ **do**
      $\mathbb{M}^{\rightarrow}_{r'} \coloneqq \{\emptyset\}$
      for every $s \in \Sigma \cap \mathsf{N_R}$ *with* $\mathcal{T}_1 \models r \sqsubseteq s$ **do**
        $\mathbb{M}^{\rightarrow}_{r'} \coloneqq \mathbb{M}^{\rightarrow}_{r'} \otimes \text{Just}_{\mathcal{T}_2}(r' \sqsubseteq s)$
      $\mathbb{M}^{\rightarrow}_{Z'} \coloneqq \textsc{Cover}_{\rightarrow}(\mathcal{T}_1, Z, \Sigma, \mathcal{T}_2, Z')$
      $\mathbb{M}^{\rightarrow}_{\exists r.Z} \coloneqq \mathbb{M}^{\rightarrow}_{\exists r.Z} \cup \big(\text{Just}_{\mathcal{T}_2}(X_2 \sqsubseteq Y') \otimes \{\{Y' \bowtie_2 \exists r'.Z'\}\} \otimes \mathbb{M}^{\rightarrow}_{r'} \otimes \mathbb{M}^{\rightarrow}_{Z'}\big)$
    $\mathbb{M}^{\rightarrow}_{(X_1,X_2)} \coloneqq \mathbb{M}^{\rightarrow}_{(X_1,X_2)} \otimes \mathbb{M}^{\rightarrow}_{\exists r.Z}$
  return $\mathbb{M}^{\rightarrow}_{(X_1,X_2)}$

Algorithm 3: Computing a Cover of all Subsumee Justifications

function $\textsc{Cover}^{\mathsf{cq}}_{\leftarrow}$ $(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2, \zeta)$
  if X_1 *is not* Σ-entailed *w.r.t.* $\mathcal{T}_1$ **then**
    return $\{\emptyset\}$
  $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \textsc{Cover}^{\mathsf{N_C}}_{\leftarrow}(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2, \zeta)$
  if X_1 *is not complex* Σ*-entailed in* $\mathcal{T}_1$ **then**
    return $\mathbb{M}^{\leftarrow}_{(X_1,X_2)}$
  if $X_1 \equiv \exists r.Y \in \mathcal{T}_1$, *and* r, Y *are* Σ*-entailed w.r.t.* $\mathcal{T}_1$ **then**
    $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \mathbb{M}^{\leftarrow}_{(X_1,X_2)} \otimes \textsc{Cover}^{\exists}_{\leftarrow}(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2, \zeta)$
  else if $X_1 \equiv Y_1 \sqcap \ldots \sqcap Y_m \in \mathcal{T}_1$ **then**
    $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \mathbb{M}^{\leftarrow}_{(X_1,X_2)} \otimes \textsc{Cover}^{\sqcap}_{\leftarrow}(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2, \zeta)$
  return $\text{Minimise}_{\subseteq}(\mathbb{M}^{\leftarrow}_{(X_1,X_2)})$

Algorithm 4: Computing a Cover of all Subsumee Projection Justifications ($S^{\leftarrow}_{\mathsf{N_C}}$)

function $\textsc{Cover}^{\mathsf{N_C}}_{\leftarrow}$ $(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2, \zeta)$
  $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \{\emptyset\}$
  for *every* $B \in \Sigma^{\zeta}$ *such that* $\mathcal{T}_1 \models B \sqsubseteq X_1$ **do**
    for *every* $X_2 \in$ *non-conj*$_{\mathcal{T}_2}(X_1)$ *such that* $\zeta = \varepsilon$ *or* $\mathcal{T}_2 \models \mathsf{ran}(\zeta) \sqsubseteq X_2$ **do**
      $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \mathbb{M}^{\leftarrow}_{(X_1,X_2)} \otimes \text{Just}_{\mathcal{T}_2}(B \sqsubseteq X_2)$
  return $\mathbb{M}^{\leftarrow}_{(X_1,X_2)}$

Algorithm 5: Computing a Cover of all Subsumee Projection Justifications ($S^{\leftarrow}_{\exists}$)

function $\textsc{Cover}^{\exists}_{\leftarrow}$ $(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2, \zeta)$
  let $\alpha_{X_1} \coloneqq X_1 \equiv \exists r.Y_1 \in \mathcal{T}_1$
  $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \{\text{max-tree}^{\sqcap}_{\mathcal{T}_2}(X_2)\}$
  for every $s \in \Sigma \cap \mathsf{N_R}$ *such that* $\mathcal{T}_1 \models s \sqsubseteq r$ **do**
    for every $X_2' \in \text{non-conj}_{\mathcal{T}_2}(X_2)$ such that $\zeta \neq \varepsilon$ implies $\mathcal{T}_2 \not\models \mathsf{ran}(\zeta) \sqsubseteq X_2'$ and $\mathcal{T}_2 \not\models \mathsf{dom}(s) \sqsubseteq X_2'$ **do**
      let $\alpha_{X_2'} \coloneqq X_2' \equiv \exists r'.Y_2' \in \mathcal{T}_2$
      $\mathbb{M}^{\leftarrow}_{Y_2'} \coloneqq \textsc{Cover}^{\mathsf{cq}}_{\leftarrow}(\mathcal{T}_1, Y_1, \Sigma, \mathcal{T}_2, Y_2', s)$
      $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \mathbb{M}^{\leftarrow}_{(X_1,X_2)} \otimes \big(\{\{\alpha_{X_2'}\}\} \otimes \text{Just}_{\mathcal{T}_2}(s \sqsubseteq r) \otimes \mathbb{M}^{\leftarrow}_{Y_2'}\big)$
  return $\mathbb{M}^{\leftarrow}_{(X_1,X_2)}$

Algorithm 6: Computing a Cover of all Subsumee Justifications ($S^{\leftarrow}_{\sqcap}$)

function $\textsc{Cover}^{\sqcap}_{\leftarrow}$ $(\mathcal{T}_1, X_1, \Sigma, \mathcal{T}_2, X_2, \zeta)$
  let $\alpha_{X_1} \coloneqq X_1 \equiv Y_1 \sqcap \ldots \sqcap Y_m \in \mathcal{T}_1$
  $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \emptyset$
  for every $\Gamma \in \text{DefForest}^{\sqcap}_{\mathcal{T}_2}(X_2)$ **do**
    let $\delta_{\Gamma} \coloneqq \{\, \text{def}^{\sqcap}_{\mathcal{T}_2}(X') \mid X' \in \text{leaves}(\Gamma) \cap \text{def}^{\sqcap}_{\mathcal{T}_2} \,\}$
    $\mathbb{M}^{\leftarrow}_{\Gamma} \coloneqq \{\Gamma\}$
    for every $X_2' \in \text{leaves}(\Gamma)$ such that $\zeta = \varepsilon$ or $\mathcal{T}_2 \not\models \mathsf{ran}(\zeta) \sqsubseteq X_2'$ **do**
      $\mathbb{M}^{\leftarrow}_{X_2'} \coloneqq \emptyset$
      for every $X_1' \in \text{non-conj}_{\mathcal{T}_1}(X_1)$ such that $\zeta = \varepsilon$ or $\mathcal{T}_2 \not\models \mathsf{ran}(\zeta) \sqsubseteq X_1'$ **do**
        if $\langle \mathcal{T}_1, X_1'\rangle \sim^{\leftarrow}_{\Sigma,\zeta} \langle \mathcal{T}_2 \setminus \delta_{\Gamma}, X_2'\rangle$ **then**
          $\mathbb{M}^{\leftarrow}_{X_2'} \coloneqq \mathbb{M}^{\leftarrow}_{X_2'} \cup \textsc{Cover}^{\mathsf{cq}}_{\leftarrow}(\mathcal{T}_1, X_1', \Sigma, \mathcal{T}_2 \setminus \delta_{\Gamma}, X_2', \epsilon)$
      $\mathbb{M}^{\leftarrow}_{\Gamma} \coloneqq \mathbb{M}^{\leftarrow}_{\Gamma} \otimes \mathbb{M}^{\leftarrow}_{X_2'}$
    $\mathbb{M}^{\leftarrow}_{(X_1,X_2)} \coloneqq \mathbb{M}^{\leftarrow}_{(X_1,X_2)} \cup \mathbb{M}^{\leftarrow}_{\Gamma}$
  return $\mathbb{M}^{\leftarrow}_{(X_1,X_2)}$

Algorithm 7: Computing a Cover of all Subsumption Justifications for Role Inclusions

function $\textsc{Cover}^{R}$ $(\mathcal{T}_1, \Sigma, \mathcal{T}_2)$
  $\mathbb{M} = \{\emptyset\}$
  for every $r, s \in \Sigma \cap \mathsf{N_R}$ *such that* $\mathcal{T}_1 \models r \sqsubseteq s$ **do**
    $\mathbb{M} \coloneqq \mathbb{M} \otimes \text{Just}_{\mathcal{T}_2}(r \sqsubseteq s)$
  return $\text{Minimise}_{\subseteq}(\mathbb{M})$

Fig. 1. Algorithms of computing all subsumer and subsumee justifications

The existence of axiom $\alpha_{X_1} \coloneqq X_1 \equiv \exists r.Y_1 \in \mathcal{T}_1$ in Line 2 of Algorithm 5 is guaranteed by Line 7 of Algorithm 3. The axiom $\alpha_{X_2'} \coloneqq X_2' \equiv \exists r'.Y_2' \in \mathcal{T}_2$ in Line 6 of Algorithm 5 exists as we assume that X_2 in $\mathcal{T}_2$ "subsumee-simulates" X_1 in $\mathcal{T}_1$. Moreover, there is at most one axiom $\alpha_{X_1} \in \mathcal{T}_1$ and at most one $\alpha_{X_2'} \in \mathcal{T}_2$ as $\mathcal{T}_1$ and $\mathcal{T}_2$ are terminologies. The concept name X_2 may be defined as a conjunction in $\mathcal{T}_2$ whose conjuncts in turn may also be defined as a conjunction in $\mathcal{T}_2$ and so forth. In Line 3 all axioms forming the maximal resulting definitorial conjunctive tree are collected.

For the next algorithm, we define $\mathrm{def}^{\sqcap}_{\mathcal{T}} \coloneqq \{ X \in \mathrm{sig}^{\mathsf{N_C}}(\mathcal{T}) \mid X \equiv Y_1 \sqcap \ldots \sqcap Y_n \in \mathcal{T} \}$ to be the set of concept names that are conjunctively defined in $\mathcal{T}$. For every $X \in \mathrm{def}^{\sqcap}_{\mathcal{T}}$, we set $\mathrm{def}^{\sqcap}_{\mathcal{T}}(X) \coloneqq \alpha$, where $\alpha = X \equiv Y_1 \sqcap \ldots \sqcap Y_n \in \mathcal{T}$.

The axiom $\alpha_{X_1} \coloneqq X_1 \equiv Y_1 \sqcap \ldots \sqcap Y_m \in \mathcal{T}_1$ in Line 2 of Algorithm 6 is guaranteed by Line 9 of Algorithm 3. In case X_2 is defined as a conjunction in $\mathcal{T}_2$, the pair consisting of $\mathcal{T}_2$ containing only a partial conjunctive tree rooted at X_2 and X_2 needs to be considered to be sufficient to "subsumee simulate" X_1 in $\mathcal{T}_1$. Therefore Algorithm 3 considers every partial conjunctive tree Γ from $\mathrm{DefForest}^{\sqcap}_{\mathcal{T}_2}(X_2)$ in Line 4 and removes the axioms in δ_Γ connecting the leaves of Γ with the remaining conjunctive tree from $\mathcal{T}_2$ in lines 10 and 11.

The following theorem shows that Algorithm 3 indeed computes a cover of the set of subsumee modules. Thus every subsumee justification is guaranteed to be among the computed sets of axioms.

Theorem 8. *Let $\chi = \langle \mathcal{T}_1, \varphi_1, \Sigma, \mathcal{T}_2, \varphi_2 \rangle$ and $\varphi_1, \varphi_2 \in (\Sigma \cap \mathsf{N_C}) \cup \Sigma^{\mathsf{dom}} \cup \Sigma^{\mathsf{ran}}$. Additionally, let $\mathbb{M} \coloneqq \textsc{Cover}^{\mathsf{cq}}_{\leftarrow}(\mathcal{T}_1, \varphi_1, \Sigma, \mathcal{T}_2, \varphi_2, \epsilon)$. If $\mathcal{T}_1 \sim^{\leftarrow}_{\Sigma} \mathcal{T}_2$, then $\mathbb{M}$ is the set of all subsumee justifications under χ.*

The number of (minimal) projection justifications depends on $\mathcal{T}_1$, $\mathcal{T}_2$ and Σ. In general, this number is bounded by an exponential in the size of the terminologies. The simulation checks can be performed in polynomial time [11, 23]. Our algorithm for computing projection justifications, therefore, runs in time exponential in the size of the input.

6 Conclusion

We introduce the notion of projection module for conjunctive queries as a subset of an ontology capturing the knowledge about a given signature as specified in a reference ontology. Here knowledge about a signature means the set of entailed conjunctive queries about the signature. This allows comparing ontologies in a more fine-grained fashion compared to merely extracting modules. Projection modules enable us to check how knowledge is implemented in terms of axioms in different ontologies. In particular, we can verify that and how specifications as defined in reference ontologies have been realised. We have presented algorithms for computing projection modules of acyclic $\mathcal{ELH}^r$-terminologies w.r.t. conjunctive queries. Similar algorithms can be used to deal with projection modules for concept subsumption queries and instance queries. We expect that the algorithms can be extended to deal with cyclic terminologies and even general $\mathcal{ELH}^r$-TBoxes.

References

1. Alsubait, T., Parsia, B., Sattler, U.: Measuring similarity in ontologies: A new family of measures. In: Proceedings of EKAW'14: the 19th International Conference on Knowledge Engineering and Knowledge Management. pp. 13–25 (2014)
2. Arif, M.F., Mencía, C., Ignatiev, A., Manthey, N., Peñaloza, R., Marques-Silva, J.: BEACON: An efficient SAT-based tool for debugging EL^{+}-ontologies. In: Proceedings of SAT'16: the 19th International Conference on the Theory and Applications of Satisfiability Testing. pp. 521–530 (2016)
3. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The description logic handbook: theory, implementation, and applications. Cambridge University Press, 2 edn. (June 2010)
4. Baader, F., Horrocks, I., Lutz, C., Sattler, U.: An Introduction to Description Logic. Cambridge University Press (2017)
5. Baader, F., Peñaloza, R., Suntisrivaraporn, B.: Pinpointing in the description logic EL. In: Proceedings of DL'07: the 20th International Workshop on Description Logics (2007)
6. Botoeva, E., Konev, B., Lutz, C., Ryzhikov, V., Wolter, F., Zakharyaschev, M.: Inseparability and conservative extensions of description logic ontologies: A survey. In: Reasoning Web Summer School 2016, LNCS, vol. 9885, pp. 27–89. Springer International Publishing (2017)
7. Botoeva, E., Kontchakov, R., Ryzhikov, V., Wolter, F., Zakharyaschev, M.: Games for query inseparability of description logic knowledge bases. Artificial Intelligence 234, 78 – 119 (2016)
8. Botoeva, E., Lutz, C., Ryzhikov, V., Wolter, F., Zakharyaschev, M.: Query-based entailment and inseparability for alc ontologies. In: Proceedings of IJCAI'16: the 25th International Joint Conference on Artificial Intelligence. pp. 1001–1007. AAAI Press (2016)
9. Chen, J., Ludwig, M., Ma, Y., Walther, D.: Zooming in on ontologies: Minimal modules and best excerpts. In: Proceedings Part I of ISWC'17: the 16th International Semantic Web Conference. pp. 173–189 (2017)
10. Domingue, J., Anutariya, C. (eds.): Proceedings of ASWC'08: the 3rd Asian Semantic Web Conference on The Semantic Web, Lecture Notes in Computer Science, vol. 5367. Springer (2008)
11. Ecke, A., Ludwig, M., Walther, D.: The concept difference for $\mathcal{EL}$-terminologies using hypergraphs. In: Proceedings of DChanges'13 (2013)
12. Euzenat, J., Shvaiko, P.: Ontology Matching, Second Edition. Springer (2013)
13. Jung, J.C., Lutz, C., Martel, M., Schneider, T.: Query conservative extensions in horn description logics with inverse roles. In: Proceedings of IJCAI'17: the 26th International Joint Conference on Artificial Intelligence. pp. 1116–1122. AAAI Press (2017)
14. Kalyanpur, A., Parsia, B., Sirin, E., Hendler, J.A.: Debugging unsatisfiable classes in OWL ontologies. Journal of Web Semantics 3(4), 268–293 (2005)
15. Kazakov, Y., Skocovsky, P.: Enumerating justifications using resolution. In: Proceedings of DL'17: the 30th International Workshop on Description Logics (2017)
16. Konev, B., Kontchakov, R., Ludwig, M., Schneider, T., Wolter, F., Zakharyaschev, M.: Conjunctive query inseparability of OWL 2 QL TBoxes. In: Proceedings of AAAI'11: the 25th Conference on Artificial Intelligence. AAAI Press (2011)
17. Konev, B., Ludwig, M., Walther, D., Wolter, F.: The logical difference for the lightweight description logic EL. Journal of Artificial Intelligence Research 44, 633–708 (2012)

18. Konev, B., Lutz, C., Walther, D., Wolter, F.: Semantic modularity and module extraction in description logics. In: Proceedings of ECAI'08: the 18th European Conference on Artificial Intelligence. pp. 55–59. IOS Press (2008)
19. Konev, B., Lutz, C., Walther, D., Wolter, F.: Model-theoretic inseparability and modularity of description logic ontologies. Artificial Intelligence 203, 66–103 (2013)
20. Konev, B., Walther, D., Wolter, F.: The logical difference problem for description logic terminologies. In: Proceedings of IJCAR'08. pp. 259–274 (2008)
21. Kontchakov, R., Wolter, F., Zakharyaschev, M.: Logic-based ontology comparison and module extraction, with an application to DL-Lite. Artificial Intelligence 174(15), 1093–1141 (2010)
22. Lehmann, K., Turhan, A.: A framework for semantic-based similarity measures for ELH-concepts. In: Proceedings of JELIA'12: the 13th European Conference Logics in Artificial Intelligence. Lecture Notes in Computer Science, vol. 7519, pp. 307–319. Springer (2012)
23. Ludwig, M., Walther, D.: The logical difference for ELHr-terminologies using hypergraphs. In: Proceedings of ECAI'14: the 21st European Conference on Artificial Intelligence. pp. 555–560 (2014)
24. Romero, A.A., Kaminski, M., Grau, B.C., Horrocks, I.: Module extraction in expressive ontology languages via datalog reasoning. Journal of Artificial Intelligence Research 55, 499–564 (2016)
25. Sattler, U., Schneider, T., Zakharyaschev, M.: Which kind of module should I extract? In: Proceedings of DL'09. CEUR Workshop Proceedings, vol. 477. CEUR-WS.org (2009)
26. Zhou, Z., Qi, G., Suntisrivaraporn, B.: A new method of finding all justifications in OWL 2 EL. In: Proceedings of WI'13: IEEE/WIC/ACM International Conferences on Web Intelligence. pp. 213–220 (2013)