

Connection-minimal Abduction in \mathcal{EL} via Translation to FOL

Fajar Haifani^{1,2}, Patrick Koopmann³, Sophie Tourret^{4,1}, and Christoph Weidenbach¹

¹ Max-Planck-Institut für Informatik, Saarland Informatics Campus, Saarbrücken, Germany

² Graduate School of Computer Science, Saarbrücken, Germany
`{f.haifani,c.weidenbach}@mpi-inf.mpg.de`

³ TU Dresden, Dresden, Germany

`patrick.koopmann@tu-dresden.de`

⁴ Université de Lorraine, CNRS, Inria, LORIA, Nancy, France
`sophie.tourret@inria.fr`

Abstract. Abduction in description logics finds extensions of a knowledge base to make it entail an observation. As such, it can be used to explain why the observation does not follow, to repair incomplete knowledge bases, and to provide possible explanations for unexpected observations. We consider TBox abduction in the lightweight description logic \mathcal{EL} , where the observation is a concept inclusion and the background knowledge is a TBox, i.e., a set of concept inclusions. To avoid useless answers, such problems usually come with further restrictions on the solution space and/or minimality criteria that help sort the chaff from the grain. We argue that existing minimality notions are insufficient, and introduce connection minimality. This criterion follows Occam’s razor by rejecting hypotheses that use concept inclusions unrelated to the problem at hand. We show how to compute a special class of connection-minimal hypotheses in a sound and complete way. Our technique is based on a translation to first-order logic, and constructs hypotheses based on prime implicates. We evaluate a prototype implementation of our approach on ontologies from the medical domain.

1 Introduction

Ontologies are used in areas like biomedicine or the semantic web to represent and reason about terminological knowledge. They consist normally of a set of axioms formulated in a description logic (DL), giving definitions of concepts, or stating relations between them. In the lightweight description logic \mathcal{EL} [2], particularly used in the biomedical domain, we find ontologies that contain around a hundred thousand axioms. For instance, SNOMED CT⁵ contains over 350,000 axioms, and the Gene Ontology GO⁶ defines over 50,000 concepts. A central reasoning task for ontologies is to determine whether one concept is subsumed

⁵ <https://www.snomed.org/> ⁶ <http://geneontology.org/>

by another, a question that can be answered in polynomial time [1], and rather efficiently in practice using highly optimized description logic reasoners [30]. If the answer to this question is unexpected or hints at an error, a natural interest is in an explanation for that answer—especially if the ontology is complex. But whereas explaining entailments—i.e., explaining why a concept subsumption holds—is well-researched in the DL literature and integrated into standard ontology editors [22, 23], the problem of explaining non-entailments has received less attention, and there is no standard tool support. Classical approaches involve counter-examples [5], or *abduction*.

In abduction a non-entailment $\mathcal{T} \not\models \alpha$, for a TBox \mathcal{T} and an observation α , is explained by providing a “missing piece”, the *hypothesis*, that, when added to the ontology, would entail α . Thus it provides possible fixes in case the entailment should hold. In the DL context, depending on the shape of the observation, one distinguishes between concept abduction [6], ABox abduction [7–10, 13, 20, 25, 26, 31, 32], TBox abduction [11, 34] or knowledge base abduction [15, 27]. We are focusing here on TBox abduction, where the ontology and hypothesis are TBoxes and the observation is a concept inclusion (CI), i.e., a single TBox axiom.

To illustrate this problem, consider the following TBox, about academia,

$$\begin{aligned} \mathcal{T}_a = \{ & \exists \text{employment.ResearchPosition} \sqcap \exists \text{qualification.Diploma} \sqsubseteq \text{Researcher}, \\ & \exists \text{writes.ResearchPaper} \sqsubseteq \text{Researcher}, \text{Doctor} \sqsubseteq \exists \text{qualification.PhD}, \\ & \text{Professor} \equiv \text{Doctor} \sqcap \exists \text{employment.Chair}, \\ & \text{FundsProvider} \sqsubseteq \exists \text{writes.GrantApplication} \} \end{aligned}$$

that states, in natural language:

- “Being employed in a research position and having a qualifying diploma implies being a researcher.”
- “Writing a research paper implies being a researcher.”
- “Being a doctor implies holding a PhD qualification.”
- “Being a professor is being a doctor employed at a (university) chair.”
- “Being a funds provider implies writing grant applications.”

The observation $\alpha_a = \text{Professor} \sqsubseteq \text{Researcher}$, “Being a professor implies being a researcher”, does not follow from \mathcal{T}_a although it should. We can use TBox abduction to find different ways of recovering this entailment.

Commonly, to avoid trivial answers, the user provides syntactic restrictions on hypotheses, such as a set of abducible axioms to pick from [8, 31], a set of abducible predicates [26, 27], or patterns on the shape of the solution [12]. But even with those restrictions in place, there may be many possible solutions and, to find the ones with the best explanatory potential, syntactic criteria are usually combined with minimality criteria such as subset minimality, size minimality, or semantic minimality [7]. Even combined, these minimality criteria still retain a major flaw. They allow for explanations that go against the principle of parsimony, also known as Occam’s razor, in that they may contain concepts

that are completely unrelated to the problem at hands. As an illustration, let us return to our academia example. The TBoxes

$$\begin{aligned} \mathcal{H}_{a1} &= \{ \text{Chair} \sqsubseteq \text{ResearchPosition}, \text{PhD} \sqsubseteq \text{Diploma} \} \text{ and} \\ \mathcal{H}_{a2} &= \{ \text{Professor} \sqsubseteq \text{FundsProvider}, \text{GrantApplication} \sqsubseteq \text{ResearchPaper} \} \end{aligned}$$

are two hypotheses solving the TBox abduction problem involving \mathcal{T}_a and α_a . Both of them are subset-minimal, have the same size, and are incomparable w.r.t. the entailment relation, so that traditional minimality criteria cannot distinguish them. However, intuitively, the second hypothesis feels more arbitrary than the first. Looking at \mathcal{H}_{a1} , *Chair* and *ResearchPosition* occur in \mathcal{T}_a in concept inclusions where the concepts in α_a also occur, and both *PhD* and *Diploma* are similarly related to α_a but via the role *qualification*. In contrast, \mathcal{H}_{a2} involves the concepts *FundsProvider* and *GrantApplication* that are not related to α_a in any way in \mathcal{T}_a . In fact, any random concept inclusion $A \sqsubseteq \exists \text{writes}.B$ in \mathcal{T}_a would lead to a hypothesis similar to \mathcal{H}_{a2} where A replaces *FundsProvider* and B replaces *GrantApplication*. Such explanations are not parsimonious.

We introduce a new minimality criterion called *connection minimality* that is parsimonious (Sect. 3), defined for the lightweight description logic \mathcal{EL} . This criterion characterizes hypotheses for \mathcal{T} and α that connect the left- and right-hand sides of the observation α without introducing spurious connections. To achieve this, every left-hand side of a CI in the hypothesis must follow from the left-hand side of α in \mathcal{T} , and, taken together, all the right-hand sides of the CIs in the hypothesis must imply the right-hand side of α in \mathcal{T} , as is the case for \mathcal{H}_{a1} . To compute connection-minimal hypotheses in practice, we present a technique based on first-order reasoning that proceeds in three steps (Sect. 4). First, we translate the abduction problem into a first-order formula Φ . We then compute the prime implicates of Φ , that is, a set of minimal logical consequences of Φ that subsume all other consequences of Φ . In the final step, we construct, based on those prime implicates, solutions to the original problem. We prove that all hypotheses generated in this way satisfy the connection minimality criterion, and that the method is complete for a relevant subclass of connection-minimal hypotheses. We use the SPASS theorem prover [35] as a restricted SOS-resolution [19, 36] engine for the computation of prime implicates in a prototype implementation (Sect. 5), and we present an experimental analysis of its performances on a set of bio-medical ontologies (Sect. 6). Our results indicate that our method can in many cases be applied in practice to compute connection-minimal hypotheses. A technical report companion of this paper includes all proofs as well as a detailed example of our method as appendices [17].

There are not many techniques that can handle TBox abduction in \mathcal{EL} or more expressive DLs [11, 27, 34]. In [11], instead of a set of abducibles, a set of *justification patterns* is given, in which the solutions have to fit. An arbitrary oracle function is used to decide whether a solution is admissible or not (which may use abducibles, justification patterns, or something else), and it is shown that deciding the existence of hypotheses is tractable. However, different to our approach, they only consider atomic CIs in hypotheses, while we also

allow for hypotheses involving conjunction. The setting from [34] also considers \mathcal{EL} , and abduction under various minimality notions such as subset minimality and size minimality. It presents practical algorithms, and an evaluation of an implementation for an always-true informativeness oracle (i.e., limited to subset minimality). Different to our approach, it uses an external DL reasoner to decide entailment relationships. In contrast, we present an approach that directly exploits first-order reasoning, and thus has the potential to be generalisable to more expressive DLs.

While dedicated resolution calculi have been used before to solve abduction in DLs [9, 27], to the best of our knowledge, the only work that relies on first-order reasoning for DL abduction is [25]. Similar to our approach, it uses SOS-resolution, but to perform ABox abduction for the more expressive DL \mathcal{ALC} . Apart from the different problem solved, in contrast to [25] we also provide a semantic characterization of the hypotheses generated by our method. We believe this characterization to be a major contribution of our paper. It provides an intuition of what parsimony is for this problem, independently of one's ease with first-order logic calculi, which should facilitate the adoption of this minimality criterion by the DL community. Thanks to this characterization, our technique is calculus agnostic. Any method to compute prime implicates in first-order logic can be a basis for our abduction technique, without additional theoretical work, which is not the case for [25]. Thus, abduction in \mathcal{EL} can benefit from the latest advances in prime implicates generation in first-order logic.

2 Preliminaries

We first recall the description logic \mathcal{EL} and its translation to first-order logic [2], as well as TBox abduction in this logic.

Let \mathbb{N}_C and \mathbb{N}_R be pair-wise disjoint, countably infinite sets of unary predicates called *atomic concepts* and of binary predicates called *roles*, respectively. Generally, we use letters A, B, E, F, \dots for atomic concepts, and r for roles, possibly annotated. Letters C, D , possibly annotated, denote \mathcal{EL} *concepts*, built according to the syntax rule

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

We implicitly represent \mathcal{EL} conjunctions as sets, that is, without order, nested conjunctions, and multiple occurrences of a conjunct. We use $\prod\{C_1, \dots, C_m\}$ to abbreviate $C_1 \sqcap \dots \sqcap C_m$, and identify the empty conjunction ($m = 0$) with \top . An \mathcal{EL} *TBox* \mathcal{T} is a finite set of *concept inclusions* (CIs) of the form $C \sqsubseteq D$.

\mathcal{EL} is a syntactic variant of a fragment of first-order logic that uses \mathbb{N}_C and \mathbb{N}_R as predicates. Specifically, TBoxes \mathcal{T} and CIs α correspond to closed first-order formulas $\pi(\mathcal{T})$ and $\pi(\alpha)$ resp., while concepts C correspond to open formulas

$\pi(C, x)$ with a free variable x . In particular, we have

$$\begin{aligned} \pi(\top, x) &:= \mathbf{true}, & \pi(\exists r.C, x) &:= \exists y.(r(x, y) \wedge \pi(C, y)), \\ \pi(A, x) &:= A(x), & \pi(C \sqsubseteq D) &:= \forall x.(\pi(C, x) \rightarrow \pi(D, x)), \\ \pi(C \sqcap D, x) &:= \pi(C, x) \wedge \pi(D, x), & \pi(\mathcal{T}) &:= \bigwedge \{\pi(\alpha) \mid \alpha \in \mathcal{T}\}. \end{aligned}$$

As common, we often omit the \bigwedge in conjunctions $\bigwedge \Phi$, that is, we identify sets of formulas with the conjunction over those. The notions of a *term* t ; an *atom* $P(\bar{t})$ where \bar{t} is a sequence of terms; a *positive literal* $P(\bar{t})$; a *negative literal* $\neg P(\bar{t})$; and a clause, Horn, definite, positive or negative, are defined as usual for first-order logic, and so are entailment and satisfaction of first-order formulas.

We identify CIs and TBoxes with their translation into first-order logic, and can thus speak of the entailment between formulas, CIs and TBoxes. When $\mathcal{T} \models C \sqsubseteq D$ for some \mathcal{T} , we call C a *subsumee* of D and D a *subsumer* of C . We adhere here to the definition of the word “subsume”: “to include or contain something else”, although the terminology is reversed in first-order logic. We say two TBoxes $\mathcal{T}_1, \mathcal{T}_2$ are *equivalent*, denoted $\mathcal{T}_1 \equiv \mathcal{T}_2$ iff $\mathcal{T}_1 \models \mathcal{T}_2$ and $\mathcal{T}_2 \models \mathcal{T}_1$. For example $\{D \sqsubseteq C_1, \dots, D \sqsubseteq C_n\} \equiv \{D \sqsubseteq C_1 \sqcap \dots \sqcap C_n\}$. It is well known that, due to the absence of concept negation, every \mathcal{EL} TBox is consistent.

The abduction problem we are concerned with in this paper is the following:

Definition 1. *An \mathcal{EL} TBox abduction problem (shortened to abduction problem) is a tuple $\langle \mathcal{T}, \Sigma, C_1 \sqsubseteq C_2 \rangle$, where \mathcal{T} is a TBox called the background knowledge, Σ is a set of atomic concepts called the abducible signature, and $C_1 \sqsubseteq C_2$ is a CI called the observation, s.t. $\mathcal{T} \not\models C_1 \sqsubseteq C_2$. A solution to this problem is a TBox*

$$\mathcal{H} \subseteq \{A_1 \sqcap \dots \sqcap A_n \sqsubseteq B_1 \sqcap \dots \sqcap B_m \mid \{A_1, \dots, A_n, B_1, \dots, B_m\} \subseteq \Sigma\}$$

where $m > 0$, $n \geq 0$ and such that $\mathcal{T} \cup \mathcal{H} \models C_1 \sqsubseteq C_2$ and, for all CIs $\alpha \in \mathcal{H}$, $\mathcal{T} \not\models \alpha$. A solution to an abduction problem is called a hypothesis.

For example, \mathcal{H}_{a1} and \mathcal{H}_{a2} are solutions for $\langle \mathcal{T}_a, \Sigma, \alpha_a \rangle$, as long as Σ contains all the atomic concepts that occur in them. Note that in our setting, as in [6, 34], concept inclusions in a hypothesis are *flat*, i.e., they contain no existential role restrictions. While this restricts the solution space for a given problem, it is possible to bypass this limitation in a targeted way, by introducing fresh atomic concepts equivalent to a concept of interest. We exclude the consistency requirement $\mathcal{T} \cup \mathcal{H} \not\models \perp$, that is given in other definitions of DL abduction problem [26], since \mathcal{EL} TBoxes are always consistent. We also allow $m > 1$ instead of the usual $m = 1$. This produces the same hypotheses modulo equivalence.

For simplicity, we assume in the following that the concepts C_1 and C_2 in the abduction problem are atomic. We can always introduce fresh atomic concepts A_1 and A_2 with $A_1 \sqsubseteq C_1$ and $C_2 \sqsubseteq A_2$ to solve the problem for complex concepts.

Common minimality criteria include *subset* minimality, *size* minimality and *semantic* minimality, that respectively favor \mathcal{H} over \mathcal{H}' if: $\mathcal{H} \subsetneq \mathcal{H}'$; the number of atomic concepts in \mathcal{H} is smaller than in \mathcal{H}' ; and if $\mathcal{H} \models \mathcal{H}'$ but $\mathcal{H}' \not\models \mathcal{H}$.

3 Connection-minimal Abduction

To address the lack of parsimony of common minimality criteria, illustrated in the academia example, we introduce *connection* minimality. Intuitively, connection minimality only accepts those hypotheses that ensure that every CI in the hypothesis is connected to both C_1 and C_2 in \mathcal{T} , as is the case for \mathcal{H}_{a1} in the academia example. The definition of connection minimality is based on the following ideas: 1) Hypotheses for the abduction problem should create a *connection* between C_1 and C_2 , which can be seen as a concept D that satisfies $\mathcal{T} \cup \mathcal{H} \models C_1 \sqsubseteq D, D \sqsubseteq C_2$. 2) To ensure parsimony, we want this connection to be based on concepts D_1 and D_2 for which we already have $\mathcal{T} \models C_1 \sqsubseteq D_1, D_2 \sqsubseteq C_2$. This prevents the introduction of unrelated concepts in the hypothesis. Note however that D_1 and D_2 can be complex, thus the connection from C_1 to D_1 (resp. D_2 to C_2) can be established by arbitrarily long chains of concept inclusions. 3) We additionally want to make sure that the connecting concepts are not more complex than necessary, and that \mathcal{H} only contains CIs that directly connect parts of D_2 to parts of D_1 by closely following their structure.

To address point 1), we simply introduce connecting concepts formally.

Definition 2. *Let C_1 and C_2 be concepts. A concept D connects C_1 to C_2 in \mathcal{T} if and only if $\mathcal{T} \models C_1 \sqsubseteq D$ and $\mathcal{T} \models D \sqsubseteq C_2$.*

Note that if $\mathcal{T} \models C_1 \sqsubseteq C_2$ then both C_1 and C_2 are connecting concepts from C_1 to C_2 , and if $\mathcal{T} \not\models C_1 \sqsubseteq C_2$, the case of interest, neither of them are.

To address point 2), we must capture *how* a hypothesis creates the connection between the concepts C_1 and C_2 . As argued above, this is established via concepts D_1 and D_2 that satisfy $\mathcal{T} \models C_1 \sqsubseteq D_1, D_2 \sqsubseteq C_2$. Note that having only two concepts D_1 and D_2 is exactly what makes the approach parsimonious. If there was only one concept, C_1 and C_2 would already be connected, and as soon as there are more than two concepts, hypotheses start becoming more arbitrary: for a very simple example with unrelated concepts, assume given a TBox that entails $\text{Lion} \sqsubseteq \text{Felidae}, \text{Mammal} \sqsubseteq \text{Animal}$ and $\text{House} \sqsubseteq \text{Building}$. A possible hypothesis to explain $\text{Lion} \sqsubseteq \text{Animal}$ is $\{\text{Felidae} \sqsubseteq \text{House}, \text{Building} \sqsubseteq \text{Mammal}\}$ but this explanation is more arbitrary than $\{\text{Felidae} \sqsubseteq \text{Mammal}\}$ —as is the case when comparing \mathcal{H}_{a2} with \mathcal{H}_{a1} in the academia example—because of the lack of connection of $\text{House} \sqsubseteq \text{Building}$ with both Lion and Animal . Clearly this CI could be replaced by any other CI entailed by \mathcal{T} , which is what we want to avoid.

We can represent the structure of D_1 and D_2 in graphs by using *EL description trees*, originally from Baader et al. [3].

Definition 3. *An EL description tree is a finite labeled tree $\mathfrak{T} = (V, E, v_0, l)$ where V is a set of nodes with root $v_0 \in V$, the nodes $v \in V$ are labeled with $l(v) \subseteq \mathbf{N}_C$, and the (directed) edges $vrw \in E$ are such that $v, w \in V$ and are labeled with $r \in \mathbf{N}_R$.*

Given a tree $\mathfrak{T} = (V, E, v_0, l)$ and $v \in V$, we denote by $\mathfrak{T}(v)$ the subtree of \mathfrak{T} that is rooted in v . If $l(v_0) = \{A_1, \dots, A_k\}$ and v_1, \dots, v_n are all the children of v_0 , we

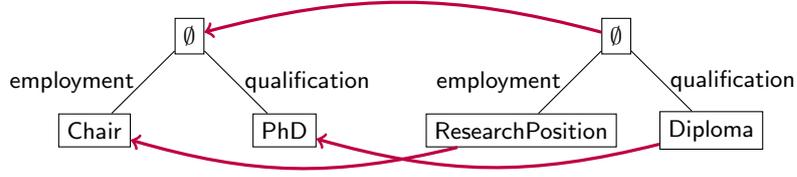


Fig. 1. Description trees of D_1 (left) and D_2 (right).

can define the concept represented by \mathfrak{T} recursively using $C_{\mathfrak{T}} = A_1 \sqcap \dots \sqcap A_k \sqcap \exists r_1.C_{\mathfrak{T}(v_1)} \sqcap \dots \sqcap \exists r_l.C_{\mathfrak{T}(v_l)}$ where for $j \in \{1, \dots, n\}$, $v_0 r_j v_j \in E$. Conversely, we can define \mathfrak{T}_C for a concept $C = A_1 \sqcap \dots \sqcap A_k \sqcap \exists r_1.C_1 \sqcap \dots \sqcap \exists r_n.C_n$ inductively based on the pairwise disjoint description trees $\mathfrak{T}_{C_i} = \{V_i, E_i, v_i, l_i\}$, $i \in \{1, \dots, n\}$. Specifically, $\mathfrak{T}_C = (V_C, E_C, v_C, l_C)$, where

$$\begin{aligned} V_C &= \{v_0\} \cup \bigcup_{i=1}^n V_i, & l_C(v) &= l_i(v) \text{ for } v \in V_i, \\ E_C &= \{v_0 r_i v_i \mid 1 \leq i \leq n\} \cup \bigcup_{i=1}^n E_i, & l_C(v_0) &= \{A_1, \dots, A_k\}. \end{aligned}$$

If $\mathcal{T} = \emptyset$, then subsumption between \mathcal{EL} concepts is characterized by the existence of a homomorphism between the corresponding description trees [3]. We generalise this notion to also take the TBox into account.

Definition 4. Let $\mathfrak{T}_1 = (V_1, E_1, v_0, l_1)$ and $\mathfrak{T}_2 = (V_2, E_2, w_0, l_2)$ be two description trees and \mathcal{T} a TBox. A mapping $\phi : V_2 \rightarrow V_1$ is a \mathcal{T} -homomorphism from \mathfrak{T}_2 to \mathfrak{T}_1 if and only if the following conditions are satisfied:

1. $\phi(w_0) = v_0$
2. $\phi(v) r \phi(w) \in E_1$ for all $vrw \in E_2$
3. for every $v \in V_1$ and $w \in V_2$ with $v = \phi(w)$, $\mathcal{T} \models \bigcap l_1(v) \sqsubseteq \bigcap l_2(w)$

If only 1 and 2 are satisfied, then ϕ is called a weak homomorphism.

\mathcal{T} -homomorphisms for a given TBox \mathcal{T} capture subsumption w.r.t. \mathcal{T} . If there exists a \mathcal{T} -homomorphism ϕ from \mathfrak{T}_2 to \mathfrak{T}_1 , then $\mathcal{T} \models C_{\mathfrak{T}_1} \sqsubseteq C_{\mathfrak{T}_2}$. This can be shown easily by structural induction using the definitions [17]. The weak homomorphism is the structure on which a \mathcal{T} -homomorphism can be built by adding some hypothesis \mathcal{H} to \mathcal{T} . It is used to reveal missing links between a subsumee D_2 of C_2 and a subsumer D_1 of C_1 , that can be added using \mathcal{H} .

Example 5. Consider the concepts

$$\begin{aligned} D_1 &= \exists \text{employment.Chair} \sqcap \exists \text{qualification.PhD} \\ D_2 &= \exists \text{employment.ResearchPosition} \sqcap \exists \text{qualification.Diploma} \end{aligned}$$

from the academia example. Figure 1 illustrates description trees for D_1 (left) and D_2 (right). The curved arrows show a weak homomorphism from \mathfrak{T}_{D_2} to

\mathfrak{T}_{D_1} that can be strengthened into a \mathcal{T} -homomorphism for some TBox \mathcal{T} that corresponds to the set of CIs in $\mathcal{H}_{a1} \cup \{\top \sqsubseteq \top\}$. The figure can also be used to illustrate what we mean by connection minimality: in order to create a connection between D_1 and D_2 , we should *only* add the CIs from $\mathcal{H}_{a1} \cup \{\top \sqsubseteq \top\}$ *unless* they are already entailed by \mathcal{T}_a . In practice, this means the weak homomorphism from D_2 to D_1 becomes a $(\mathcal{T}_a \cup \mathcal{H}_{a1})$ -homomorphism.

To address point 3), we define a partial order \preceq_{\sqcap} on concepts, s.t. $C \preceq_{\sqcap} D$ if we can turn D into C by removing conjuncts in subexpressions, e.g., $\exists r'.B \preceq_{\sqcap} \exists r.A \sqcap \exists r'.(B \sqcap B')$. Formally, this is achieved by the following definition.

Definition 6. *Let C and D be arbitrary concepts. Then $C \preceq_{\sqcap} D$ if either:*

- $C = D$,
- $D = D' \sqcap D''$, and $C \preceq_{\sqcap} D'$, or
- $C = \exists r.C'$, $D = \exists r.D'$ and $C' \preceq_{\sqcap} D'$.

We can finally capture our ideas on connection minimality formally.

Definition 7 (Connection-Minimal Abduction). *Given an abduction problem $\langle \mathcal{T}, \Sigma, C_1 \sqsubseteq C_2 \rangle$, a hypothesis \mathcal{H} is connection-minimal if there exist concepts D_1 and D_2 built over $\Sigma \cup \mathbf{N}_R$ and a mapping ϕ satisfying each of the following conditions:*

1. $\mathcal{T} \models C_1 \sqsubseteq D_1$,
2. D_2 is a \preceq_{\sqcap} -minimal concept s.t. $\mathcal{T} \models D_2 \sqsubseteq C_2$,
3. ϕ is a weak homomorphism from the tree $\mathfrak{T}_{D_2} = (V_2, E_2, w_0, l_2)$ to the tree $\mathfrak{T}_{D_1} = (V_1, E_1, v_0, l_1)$, and
4. $\mathcal{H} = \{\prod l_1(\phi(w)) \sqsubseteq \prod l_2(w) \mid w \in V_2 \wedge \mathcal{T} \not\models \prod l_1(\phi(w)) \sqsubseteq \prod l_2(w)\}$.

\mathcal{H} is additionally called packed if the left-hand sides of the CIs in \mathcal{H} cannot hold more conjuncts than they do, which is formally stated as: for \mathcal{H} , there is no \mathcal{H}' defined from the same D_2 and a D'_1 and ϕ' s.t. there is a node $w \in V_2$ for which $l_1(\phi(w)) \subsetneq l'_1(\phi'(w))$ and $l_1(\phi(w')) = l'_1(\phi'(w'))$ for $w' \neq w$.

Straightforward consequences of Def. 7 include that ϕ is a $(\mathcal{T} \cup \mathcal{H})$ -homomorphism from \mathfrak{T}_{D_2} to \mathfrak{T}_{D_1} and that D_1 and D_2 are connecting concepts from C_1 to C_2 in $\mathcal{T} \cup \mathcal{H}$ so that $\mathcal{T} \cup \mathcal{H} \models C_1 \sqsubseteq C_2$ as wanted [17]. With the help of Fig. 1 and Ex. 5, one easily establishes that hypothesis \mathcal{H}_{a1} is connection-minimal—and even packed. Connection-minimality rejects \mathcal{H}_{a2} , as a single \mathcal{T}' -homomorphism for some \mathcal{T}' between two concepts D_1 and D_2 would be insufficient: we would need two weak homomorphisms, one linking Professor to FundsProvider and another linking \exists writes.GrantApplication to \exists writes.ResearchPaper.

4 Computing Connection-minimal Hypotheses using Prime Implicates

To compute connection-minimal hypotheses in practice, we propose a method based on first-order prime implicates, that can be derived by resolution. We

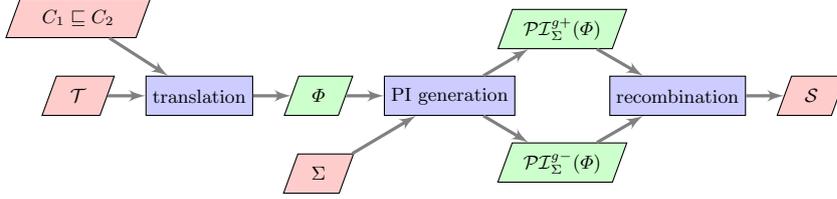


Fig. 2. \mathcal{EL} abduction using prime implicate generation in FOL.

assume the reader is familiar with the basics of first-order resolution, and do not reintroduce notions of clauses, Skolemization and resolution inferences here (for details, see [4]). In our context, every term is built on variables, denoted x, y , a single constant \mathbf{sk}_0 and unary Skolem functions usually denoted \mathbf{sk} , possibly annotated. Prime implicates are defined as follows.

Definition 8 (Prime Implicate). *Let Φ be a set of clauses. A clause φ is an implicate of Φ if $\Phi \models \varphi$. Moreover φ is prime if for any other implicate φ' of Φ s.t. $\varphi' \models \varphi$, it also holds that $\varphi \models \varphi'$.*

Let $\Sigma \subseteq \mathbf{N}_C$ be a set of unary predicates. Then $\mathcal{PI}_\Sigma^{g+}(\Phi)$ denotes the set of all positive ground prime implicates of Φ that only use predicate symbols from $\Sigma \cup \mathbf{N}_R$, while $\mathcal{PI}_\Sigma^{g-}(\Phi)$ denotes the set of all negative ground prime implicates of Φ that only use predicate symbols from $\Sigma \cup \mathbf{N}_R$.

Example 9. Given a set of clauses $\Phi = \{A_1(\mathbf{sk}_0), \neg B_1(\mathbf{sk}_0), \neg A_1(x) \vee r(x, \mathbf{sk}(x)), \neg A_1(x) \vee A_2(\mathbf{sk}(x)), \neg B_2(x) \vee \neg r(x, y) \vee \neg B_3(y) \vee B_1(x)\}$, the ground prime implicates of Φ for $\Sigma = \mathbf{N}_C$ are, on the positive side, $\mathcal{PI}_\Sigma^{g+}(\Phi) = \{A_1(\mathbf{sk}_0), A_2(\mathbf{sk}(\mathbf{sk}_0)), r(\mathbf{sk}_0, \mathbf{sk}(\mathbf{sk}_0))\}$ and, on the negative side, $\mathcal{PI}_\Sigma^{g-}(\Phi) = \{\neg B_1(\mathbf{sk}_0), \neg B_2(\mathbf{sk}_0) \vee \neg B_3(\mathbf{sk}(\mathbf{sk}_0))\}$. They are implicates because all of them are entailed by Φ . For a ground implicate φ , another ground implicate φ' such that $\varphi' \models \varphi$ and $\varphi \not\models \varphi'$ can only be obtained from φ by dropping literals. Such an operation does not produce another implicate for any of the clauses presented above as belonging to $\mathcal{PI}_\Sigma^{g+}(\Phi)$ and $\mathcal{PI}_\Sigma^{g-}(\Phi)$, thus they really are all prime.

To generate hypotheses, we translate the abduction problem into a set of first-order clauses, from which we can infer prime implicates that we then combine to obtain the result as illustrated in Fig. 2. In more details: We first translate the problem into a set Φ of Horn clauses. Prime implicates can be computed using an off-the-shelf tool [14, 29] or, in our case, a slight extension of the resolution-based version of the SPASS theorem prover [35] using the set-of-support strategy and some added features described in Sect. 5. Since Φ is Horn, $\mathcal{PI}_\Sigma^{g+}(\Phi)$ contains only unit clauses. A final recombination step looks at the clauses in $\mathcal{PI}_\Sigma^{g-}(\Phi)$ one after the other. These correspond to candidates for the connecting concepts D_2 of Def. 7. Recombination attempts to match each literal in one such clause with unit clauses from $\mathcal{PI}_\Sigma^{g+}(\Phi)$. If such a match is possible, it produces a suitable

D_1 to match D_2 , and allows the creation of a solution to the abduction problem. The set \mathcal{S} contains all the hypotheses thus obtained.

In what follows, we present our translation of abduction problems into first-order logic and formalize the construction of hypotheses from the prime implicates of this translation. We then show how to obtain termination for the prime implicate generation process with soundness and completeness guarantees on the solutions computed.

Abduction Method. We assume the \mathcal{EL} TBox in the input is in normal form as defined, e.g., by Baader et al. [2]. Thus every CI is of one of the following forms:

$$A \sqsubseteq B \quad A_1 \sqcap A_2 \sqsubseteq B \quad \exists r.A \sqsubseteq B \quad A \sqsubseteq \exists r.B$$

where $A, A_1, A_2, B \in \mathbf{N}_C \cup \{\top\}$.

The use of normalization is justified by the following lemma.

Lemma 10. *For every \mathcal{EL} TBox \mathcal{T} , we can compute in polynomial time an \mathcal{EL} TBox \mathcal{T}' in normal form such that for every other TBox \mathcal{H} and every CI $C \sqsubseteq D$ that use only names occurring in \mathcal{T} , we have $\mathcal{T} \cup \mathcal{H} \models C \sqsubseteq D$ iff $\mathcal{T}' \cup \mathcal{H} \models C \sqsubseteq D$.*

After the normalisation, we eliminate occurrences of \top , replacing this concept everywhere by the fresh atomic concept A_\top . We furthermore add $\exists r.A_\top \sqsubseteq A_\top$ and $B \sqsubseteq A_\top$ in \mathcal{T} for every role r and atomic concept B occurring in \mathcal{T} . This simulates the semantics of \top for A_\top , namely the implicit property that $C \sqsubseteq \top$ holds for any C no matter what the TBox is. In particular, this ensures that whenever there is a positive prime implicate $B(t)$ or $r(t, t')$, $A_\top(t)$ also becomes a prime implicate. Note that normalisation and \top elimination extend the signature, and thus potentially the solution space of the abduction problem. This is remedied by intersecting the set of abducible predicates Σ with the signature of the original input ontology. We assume that \mathcal{T} is in normal form and without \top in the rest of the paper.

We denote by \mathcal{T}^- the result of renaming all atomic concepts A in \mathcal{T} using fresh *duplicate* symbols A^- . This renaming is done only on concepts but not on roles, and on C_2 but not on C_1 in the observation. This ensures that the literals in a clause of $\mathcal{PI}_\Sigma^{g-}(\Phi)$ all relate to the conjuncts of a \preceq_{\top} -minimal subsumee of C_2 . Without it, some of these conjuncts would not appear in the negative implicates due to the presence of their positive counterparts as atoms in $\mathcal{PI}_\Sigma^{g+}(\Phi)$. The translation of the abduction problem $\langle \mathcal{T}, \Sigma, C_1 \sqsubseteq C_2 \rangle$ is defined as the Skolemization of

$$\pi(\mathcal{T} \uplus \mathcal{T}^-) \wedge \neg\pi(C_1 \sqsubseteq C_2^-)$$

where \mathbf{sk}_0 is used as the unique fresh Skolem constant such that the Skolemization of $\neg\pi(C_1 \sqsubseteq C_2^-)$ results in $\{C_1(\mathbf{sk}_0), \neg C_2^-(\mathbf{sk}_0)\}$. This translation is usually denoted Φ and always considered in clausal normal form.

Theorem 11. *Let $\langle \mathcal{T}, \Sigma, C_1 \sqsubseteq C_2 \rangle$ be an abduction problem and Φ be its first-order translation. Then, a TBox \mathcal{H}' is a packed connection-minimal solution to the problem if and only if an equivalent hypothesis \mathcal{H} can be constructed from non-empty sets \mathcal{A} and \mathcal{B} of atoms verifying:*

- $\mathcal{B} = \{B_1(t_1), \dots, B_m(t_m)\}$ s.t. $(\neg B_1^-(t_1) \vee \dots \vee \neg B_m^-(t_m)) \in \mathcal{PI}_\Sigma^{g^-}(\Phi)$,
- for all $t \in \{t_1, \dots, t_m\}$ there exists an A s.t. $A(t) \in \mathcal{PI}_\Sigma^{g^+}(\Phi)$,
- $\mathcal{A} = \{A(t) \in \mathcal{PI}_\Sigma^{g^+}(\Phi) \mid t \text{ is one of } t_1, \dots, t_m\}$, and
- $\mathcal{H} = \{C_{\mathcal{A},t} \sqsubseteq C_{\mathcal{B},t} \mid t \text{ is one of } t_1, \dots, t_m \text{ and } C_{\mathcal{B},t} \not\sqsubseteq_{\square} C_{\mathcal{A},t}\}$, where $C_{\mathcal{A},t} = \prod_{A(t) \in \mathcal{A}} A$ and $C_{\mathcal{B},t} = \prod_{B(t) \in \mathcal{B}} B$.

We call the hypotheses that are constructed as in Th. 11 *constructible*. This theorem states that every packed connection-minimal hypothesis is equivalent to a constructible hypothesis and vice versa. A constructible hypothesis is built from the concepts in *one* negative prime implicate in $\mathcal{PI}_\Sigma^{g^-}(\Phi)$ and *all* matching concepts from prime implicates in $\mathcal{PI}_\Sigma^{g^+}(\Phi)$. The matching itself is determined by the Skolem terms that occur in all these clauses. The subterm relation between the terms of the clauses in $\mathcal{PI}_\Sigma^{g^+}(\Phi)$ and $\mathcal{PI}_\Sigma^{g^-}(\Phi)$ is the same as the ancestor relation in the description trees of subsumers of C_1 and subsumees of C_2 respectively. The terms matching in positive and negative prime implicates allow us to identify where the missing entailments between a subsumer D_1 of C_1 and a subsumee D_2 of C_2 are. These missing entailments become the constructible \mathcal{H} . The condition $C_{\mathcal{B},t} \not\sqsubseteq_{\square} C_{\mathcal{A},t}$ is a way to write that $C_{\mathcal{A},t} \sqsubseteq C_{\mathcal{B},t}$ is not a tautology, which can be tested by subset inclusion.

The formal proof of this result is detailed in the technical report [17]. We sketch it briefly here. To start, we link the subsumers of C_1 with $\mathcal{PI}_\Sigma^{g^+}(\Phi)$. This is done at the semantics level: We show that all Herbrand models of Φ , i.e., models built on the symbols in Φ , are also models of $\mathcal{PI}_\Sigma^{g^+}(\Phi)$, that is itself such a model. Then we show that $C_1(\mathbf{sk}_0)$ as well as the formulas corresponding to the subsumers of C_1 in our translation are satisfied by all Herbrand models. This follows from the fact that Φ is in fact a set of Horn clauses. Next, we show, using a similar technique, how duplicate negative ground implicates, not necessarily prime, relate to subsumees of C_2 , with the restriction that there must exist a weak homomorphism from a description tree of a subsumer of C_1 to a description tree of the considered subsumee of C_2 . Thus, \mathcal{H} provides the missing CIs that will turn the weak homomorphism into a $(\mathcal{T} \cup \mathcal{H})$ -homomorphism. Then, we establish an equivalence between the \sqsubseteq_{\square} -minimality of the subsumee of C_2 and the primality of the corresponding negative implicate. Packability is the last aspect we deal with, whose use is purely limited to the reconstruction. It holds because \mathcal{A} contains all $A(t) \in \mathcal{PI}_\Sigma^{g^+}(\Phi)$ for all terms t occurring in \mathcal{B} .

Example 12. Consider the abduction problem $\langle \mathcal{T}_a, \Sigma, \alpha_a \rangle$ where Σ contains all concepts from \mathcal{T}_a . For the translation Φ of this problem, we have

$$\begin{aligned} \mathcal{PI}_\Sigma^{g^+}(\Phi) &= \{ \text{Professor}(\mathbf{sk}_0), \text{Doctor}(\mathbf{sk}_0), \text{Chair}(\mathbf{sk}_1(\mathbf{sk}_0)), \text{PhD}(\mathbf{sk}_2(\mathbf{sk}_0)) \} \\ \mathcal{PI}_\Sigma^{g^-}(\Phi) &= \{ \neg \text{Researcher}^-(\mathbf{sk}_0), \\ &\quad \neg \text{ResearchPosition}^-(\mathbf{sk}_1(\mathbf{sk}_0)) \vee \neg \text{Diploma}^-(\mathbf{sk}_2(\mathbf{sk}_0)) \} \end{aligned}$$

where \mathbf{sk}_1 is the Skolem function introduced for $\text{Professor} \sqsubseteq \exists \text{employment.Chair}$ and \mathbf{sk}_2 is introduced for $\text{Doctor} \sqsubseteq \exists \text{qualification.PhD}$. This leads to two constructible solutions: $\{\text{Professor} \sqcap \text{Doctor} \sqsubseteq \text{Researcher}\}$ and \mathcal{H}_{a1} , that are both

packed connection-minimal hypotheses if $\Sigma = \text{Nc}$. Another example is presented in full details in the technical report [17].

Termination. If \mathcal{T} contains cycles, there can be infinitely many prime implicates. For example, for $\mathcal{T} = \{C_1 \sqsubseteq A, A \sqsubseteq \exists r.A, \exists r.B \sqsubseteq B, B \sqsubseteq C_2\}$ both the positive and negative ground prime implicates of Φ are unbounded even though the set of constructible hypotheses is finite (as it is for any abduction problem):

$$\begin{aligned} \mathcal{PI}_{\Sigma}^{g+}(\Phi) &= \{C_1(\mathbf{sk}_0), A(\mathbf{sk}_0), A(\mathbf{sk}(\mathbf{sk}_0)), A(\mathbf{sk}(\mathbf{sk}(\mathbf{sk}_0))), \dots\}, \\ \mathcal{PI}_{\Sigma}^{g-}(\Phi) &= \{\neg C_2^-(\mathbf{sk}_0), \neg B^-(\mathbf{sk}_0), \neg B^-(\mathbf{sk}(\mathbf{sk}_0)), \dots\}. \end{aligned}$$

To find all constructible hypotheses of an abduction problem, an approach that simply computes all prime implicates of Φ , e.g., using the standard resolution calculus, will never terminate on cyclic problems. However, if we look only for subset-minimal constructible hypotheses, termination can be achieved for cyclic and non-cyclic problems alike, because it is possible to construct all such hypotheses from prime implicates that have a polynomially bounded term depth, as shown below. To obtain this bound, we consider resolution derivations of the ground prime implicates and we show that they can be done under some restrictions that imply this bound.

Before performing resolution, we compute the *presaturation* Φ_p of the set of clauses Φ , defined as

$$\Phi_p = \Phi \cup \{\neg A(x) \vee B(x) \mid \Phi \models \neg A(x) \vee B(x)\}$$

where A and B are either both original or both duplicate atomic concepts. The presaturation can be efficiently computed before the translation, using a modern \mathcal{EL} reasoner such as ELK [24], which is highly optimized towards the computation of all entailments of the form $A \sqsubseteq B$. While the presaturation computes nothing a resolution procedure could not derive, it is what allows us to bind the maximal depth of terms in inferences to that in prime implicates. If Φ_p is presaturated, we do not need to perform inferences that produce Skolem terms of a higher nesting depth than what is needed for the prime implicates.

Starting from the presaturated set Φ_p , we can show that all the relevant prime implicates can be computed if we restrict all inferences to those where

- R1** at least one premise contains a ground term,
- R2** the resolvent contains at most one variable, and
- R3** every literal in the resolvent contains Skolem terms of nesting depth at most $n \times m$, where n is the number of atomic concepts in Φ , and m is the number of occurrences of existential role restrictions in \mathcal{T} .

The first restriction turns the derivation of $\mathcal{PI}_{\Sigma}^{g+}(\Phi)$ and $\mathcal{PI}_{\Sigma}^{g-}(\Phi)$ into an SOS-resolution derivation [19] with set of support $\{C_1(\mathbf{sk}_0), C_2^-(\mathbf{sk}_0)\}$, i.e., the only two clauses with ground terms in Φ . This restriction is a straightforward consequence of our interest in computing only ground implicates, and of the fact that the non-ground clauses in Φ cannot entail the empty clause since every \mathcal{EL} TBox is consistent. The other restrictions are consequences of the following theorems, whose proofs are available in the technical report [17].

Theorem 13. *Given an abduction problem and its translation Φ , every constructible hypothesis can be built from prime implicates that are inferred under restriction **R2**.*

In fact, for $\mathcal{PT}_{\Sigma}^{g+}(\Phi)$ it is even possible to restrict inferences to generating only ground resolvents, as can be seen in the proof of Th. 13, that directly looks at the kinds of clauses that are derivable by resolution from Φ .

Theorem 14. *Given an abduction problem and its translation Φ , every subset-minimal constructible hypothesis can be built from prime implicates that have a nesting depth of at most $n \times m$, where n is the number of atomic concepts in Φ , and m is the number of occurrences of existential role restrictions in \mathcal{T} .*

The proof of Th. 14 is based on a structure called a *solution tree*, which resembles a description tree, but with multiple labeling functions. It assigns to each node a Skolem term, a set of atomic concepts called *positive label*, and a single atomic concept called *negative label*. The nodes correspond to matching partners in a constructible hypothesis: The Skolem term is the term on which we match literals. The positive label collects the atomic concepts in the positive prime implicates containing that term. The maximal anti-chains of the tree, i.e., the maximal subsets of nodes s.t. no node is the ancestor of another are such that their negative labels correspond to the literals in a derivable negative implicate. For every solution tree, the Skolem labels and negative labels of the leaves determine a negative prime implicate, and by combining the positive and negative labels of these leaves, we obtain a constructible hypothesis, called the *solution* of the tree. We show that from every solution tree with solution \mathcal{H} we can obtain a solution tree with solution $\mathcal{H}' \subseteq \mathcal{H}$ s.t. on no path, there are two nodes that agree both on the head of their Skolem labeling and on the negative label. Furthermore the number of head functions of Skolem labels is bounded by the total number n of Skolem functions, while the number of distinct negative labels is bounded by the number m of atomic concepts, bounding the depth of the solution tree for \mathcal{H}' at $n \times m$. This justifies the bound in Th 14. This bound is rather loose. For the academia example, it is equal to $22 \times 6 = 132$.

5 Implementation

We implemented our method to compute all subset-minimal constructible hypotheses in the tool CAPI.⁷ To compute the prime implicates, we used SPASS [35], a first-order theorem prover that includes resolution among other calculi. We implemented everything before and after the prime implicate computation in Java, including the parsing of ontologies, preprocessing (detailed below), clausification of the abduction problems, translation to SPASS input, as well as the parsing and processing of the output of SPASS to build the constructible hypotheses and filter out the non-subset-minimal ones. On the Java side, we used the OWL API for all DL-related functionalities [21], and the \mathcal{EL} reasoner ELK for computing the presaturations [24].

⁷ available under <https://lat.inf.tu-dresden.de/~koopmann/CAPI>

Preprocessing. Since realistic TBoxes can be too large to be processed by SPASS, we replace the background knowledge in the abduction problem by a subset of axioms relevant to the abduction problem. Specifically, we replace the abduction problem $(\mathcal{T}, \Sigma, C_1 \sqsubseteq C_2)$ by the abduction problem $(\mathcal{M}_{C_1}^\perp \cup \mathcal{M}_{C_2}^\top, \Sigma, C_1 \sqsubseteq C_2)$, where $\mathcal{M}_{C_1}^\perp$ is the \perp -module of \mathcal{T} for the signature of C_1 , and $\mathcal{M}_{C_2}^\top$ is the \top -module of \mathcal{T} for the signature of C_2 [16]. Those notions are explained in the technical report [17]. Their relevant properties are that $\mathcal{M}_{C_1}^\perp$ is a subset of \mathcal{T} s.t. $\mathcal{M}_{C_1}^\perp \models C_1 \sqsubseteq D$ iff $\mathcal{T} \models C_1 \sqsubseteq D$ for all concepts D , while $\mathcal{M}_{C_2}^\top$ is a subset of \mathcal{T} that ensures $\mathcal{M}_{C_2}^\top \models D \sqsubseteq C_2$ iff $\mathcal{T} \models D \sqsubseteq C_2$ for all concepts D . It immediately follows that every connection-minimal hypothesis for the original problem $(\mathcal{T}, \Sigma, C_1 \sqsubseteq C_2)$ is also a connection-minimal hypothesis for $(\mathcal{M}_{C_1}^\perp \cup \mathcal{M}_{C_2}^\top, \Sigma, C_1 \sqsubseteq C_2)$. For the presaturation, we compute with ELK all CIs of the form $A \sqsubseteq B$ s.t. $\mathcal{M}_{C_1}^\perp \cup \mathcal{M}_{C_2}^\top \models A \sqsubseteq B$.

Prime implicates generation. We rely on a slightly modified version of SPASS v3.9 to compute all ground prime implicates. In particular, we added the possibility to limit the number of variables allowed in the resolvents to enforce **R2**. For each of the restrictions **R1** – **R3** there is a corresponding flag (or set of flags) that is passed to SPASS as an argument.

Recombination. The construction of hypotheses from the prime implicates found in the previous stage starts with a straightforward process of matching negative prime implicates with a set of positive ones based on their Skolem terms. It is followed by subset minimality tests to discard non-subset-minimal hypotheses, since, with the bound we enforce, there is no guarantee that these are valid constructible hypotheses because the negative ground implicates they are built upon may not be prime. If SPASS terminates due to a timeout instead of reaching the bound, then it is possible that some subset-minimal constructible hypotheses are not found, and thus, some non-constructible hypotheses may be kept. Note that these are in any case solutions to the abduction problem.

6 Experiments

There is no benchmark suite dedicated to TBox abduction in \mathcal{EL} , so we created our own, using realistic ontologies from the bio-medical domain. For this, we used ontologies from the 2017 snapshot of Bioportal [28]. We restricted each ontology to its \mathcal{EL} fragment by filtering out unsupported axioms, where we replaced domain axioms and n-ary equivalence axioms in the usual way [2]. Note that, even if the ontology contains more expressive axioms, an \mathcal{EL} hypothesis is still useful if found. From the resulting set of TBoxes, we selected those containing at least 1 and at most 50,000 axioms, resulting in a set of 387 \mathcal{EL} TBoxes. Precisely, they contained between 2 and 46,429 axioms, for an average of 3,039 and a median of 569. Towards obtaining realistic benchmarks, we created three different categories of abduction problems for each ontology \mathcal{T} , where in each case, we used the signature of the entire ontology for Σ .

- Problems in **ORIGIN** use \mathcal{T} as background knowledge, and as observation a randomly chosen $A \sqsubseteq B$ s.t. A and B are in the signature of \mathcal{T} and $\mathcal{T} \not\models A \sqsubseteq B$. This covers the basic requirements of an abduction problem, but has the disadvantage that A and B can be completely unrelated in \mathcal{T} .
- Problems in **JUSTIF** contain as observation a randomly selected CI α s.t., for the original TBox, $\mathcal{T} \models \alpha$ and $\alpha \notin \mathcal{T}$. The background knowledge used is a *justification for α in \mathcal{T}* [33], that is, a minimal subset $\mathcal{I} \subseteq \mathcal{T}$ s.t. $\mathcal{I} \models \alpha$, from which a randomly selected axiom is removed. The TBox is thus a smaller set of axioms extracted from a real ontology for which we know there is a way of producing the required entailment without adding it explicitly. Justifications were computed using functionalities of the OWL API and ELK.
- Problems in **REPAIR** contain as observation a randomly selected CI α s.t. $\mathcal{T} \models \alpha$, and as background knowledge a *repair for α in \mathcal{T}* , which is a maximal subset $\mathcal{R} \subseteq \mathcal{T}$ s.t. $\mathcal{R} \not\models \alpha$. Repairs were computed using a justification-based algorithm [33] with justifications computed as for **JUSTIF**. This usually resulted in much larger TBoxes, where more axioms would be needed to establish the entailment.

All experiments were run on Debian Linux (Intel Core i5-4590, 3.30 GHz, 23 GB Java heap size). The code and scripts used in the experiments are available online [18]. The three phases of the method (see Fig. 2) were each assigned a hard time limit of 90 seconds.

For each ontology, we attempted to create and translate 5 abduction problems of each category. This failed on some ontologies because either there was no corresponding entailment (25/28/25 failures out of the 387 ontologies for **ORIGIN/JUSTIF/REPAIR**), there was a timeout during the translation (5/5/5 failures for **ORIGIN/JUSTIF/REPAIR**), or because the computation of justifications caused an exception (-/2/0 failures for **ORIGIN/JUSTIF/REPAIR**). The final number of abduction problems for each category is in the first column of Table 1.

We then attempted to compute prime implicates for these benchmarks using SPASS. In addition to the hard time limit, we gave a soft time limit of 30 seconds to SPASS, after which it should stop exploring the search space and return the implicates already found. In Table 1 we show, for each category, the percentage of problems on which SPASS succeeded in computing a non-empty set of clauses (Success) and the percentage of problems on which SPASS terminated within the time limit, where all solutions are computed (Compl.). The high number of CIs in the background knowledge explains most of the cases where SPASS reached the soft time limit. In a lot of these cases, the bound on the term depth goes into the billion, rendering it useless in practice. However, the ‘‘Compl.’’ column shows that the bound is reached before the soft time limit in most cases.

The reconstruction never reached the hard time limit. We measured the median, average and maximal number of solutions found ($\#\mathcal{H}$), size of solutions in number of CIs ($|\mathcal{H}|$), size of CIs from solutions in number of atomic concepts ($|\alpha|$), and SPASS runtime (time, in seconds), all reported in Table 1. Except for the simple **JUSTIF** problems, the number of solutions may become very large. At the same time, solutions always contain very few axioms

	#Probl.	Success	Compl.	median / avg / max			
				# \mathcal{H}	$ \mathcal{H} $	$ \alpha $	time (s.)
ORIGIN	1,925	94.7%	61.3%	1/8.51/1850	1/1.00/2	6/7.48/91	0.2/12.4/43.8
JUSTIF	1,803	100.0%	97.2%	1/1.50/5	1/1/1	2/4.21/32	0.2/1.1/34.1
REPAIR	1,805	92.9%	57.0%	43/228.05/6317	1/1.00/2	5/5.09/49	0.6/13.6/59.9

Table 1. Evaluation results.

(never more than 3), though the axioms become large too. We also noticed that highly nested Skolem terms rarely lead to more hypotheses being found: 8/1/15 for ORIGIN/JUSTIF/REPAIR, and the largest nesting depth used was: 3/1/2 for ORIGIN/JUSTIF/REPAIR. This hints at the fact that longer time limits would not have produced more solutions, and motivates future research into redundancy criteria to stop derivations (much) earlier.

7 Conclusion

We have introduced connection-minimal TBox abduction for \mathcal{EL} which finds parsimonious hypotheses, ruling out the ones that entail the observation in an arbitrary fashion. We have established a formal link between the generation of connection-minimal hypotheses in \mathcal{EL} and the generation of prime implicates of a translation Φ of the problem to first-order logic. In addition to obtaining these theoretical results, we developed a prototype for the computation of subset-minimal constructible hypotheses, a subclass of connection-minimal hypotheses that is easy to construct from the prime implicates of Φ . Our prototype uses the SPASS theorem prover as an SOS-resolution engine to generate the needed implicates. We tested this tool on a set of realistic medical ontologies, and the results indicate that the cost of computing connection-minimal hypotheses is high but not prohibitive.

We see several ways to improve our technique. The bound we computed to ensure termination could be advantageously replaced by a redundancy criterion discarding irrelevant implicates long before it is reached, thus greatly speeding computation in SPASS. We believe it should also be possible to further constrain inferences, e.g., to have them produce ground clauses only, or to generate the prime implicates with terms of increasing depth in a controlled incremental way instead of enforcing the soft time limit, but these two ideas remain to be proved feasible. As an alternative to using prime implicates, one may investigate direct method for computing connection-minimal hypotheses in \mathcal{EL} .

The theoretical worst-case complexity of connection-minimal abduction is another open question. Our method only gives a very high upper bound: by bounding only the nesting dept of Skolem terms polynomially as we did with Th. 13, we may still permit clauses with exponentially many literals, and thus double exponentially many clauses in the worst case, which would give us an 2EXPTIME upper bound to the problem of computing all subset-minimal con-

structible hypotheses. Using structure-sharing and guessing, it is likely possible to get a lower bound. We have not looked yet at lower bounds for the complexity either.

While this work focuses on abduction problems where the observation is a CI, we believe that our technique can be generalised to knowledge that also contains ground facts (ABoxes), and to observations that are of the form of conjunctive queries on the ABoxes in such knowledge bases. The motivation for such an extension is to understand why a particular query does not return any results, and to compute a set of TBox axioms that fix this problem. Since our translation already transforms the observation into ground facts, it should be possible to extend it to this setting. We would also like to generalize TBox abduction by finding a reasonable way to allow role restrictions in the hypotheses, and to extend connection-minimality to more expressive DLs such as \mathcal{ALC} .

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG), Grant 389792660 within TRR 248.

References

- [1] Baader, F., Brandt, S., Lutz, C.: Pushing the \mathcal{EL} envelope. In: Kaelbling, L.P., Saffiotti, A. (eds.) IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005. pp. 364–369. Professional Book Center (2005), <http://ijcai.org/Proceedings/05/Papers/0372.pdf>
- [2] Baader, F., Horrocks, I., Lutz, C., Sattler, U.: An Introduction to Description Logic. Cambridge University Press (2017). <https://doi.org/10.1017/9781139025355>
- [3] Baader, F., Küsters, R., Molitor, R.: Computing least common subsumers in description logics with existential restrictions. In: Proceedings of IJCAI 1999. pp. 96–103. Morgan Kaufmann (1999)
- [4] Bachmair, L., Ganzinger, H.: Resolution theorem proving. In: Robinson, J.A., Voronkov, A. (eds.) Handbook of Automated Reasoning (in 2 volumes), pp. 19–99. Elsevier and MIT Press (2001). <https://doi.org/10.1016/b978-044450813-3/50004-7>, <https://doi.org/10.1016/b978-044450813-3/50004-7>
- [5] Bauer, J., Sattler, U., Parsia, B.: Explaining by example: Model exploration for ontology comprehension. In: Grau, B.C., Horrocks, I., Motik, B., Sattler, U. (eds.) Proceedings of the 22nd International Workshop on Description Logics (DL 2009), Oxford, UK, July 27-30, 2009. CEUR Workshop Proceedings, vol. 477. CEUR-WS.org (2009), http://ceur-ws.org/Vol-477/paper_37.pdf
- [6] Bienvenu, M.: Complexity of abduction in the \mathcal{EL} family of lightweight description logics. In: Proceedings of KR 2008. pp. 220–230. AAAI Press (2008), <http://www.aaai.org/Library/KR/2008/kr08-022.php>
- [7] Calvanese, D., Ortiz, M., Simkus, M., Stefanoni, G.: Reasoning about explanations for negative query answers in DL-Lite. J. Artif. Intell. Res. **48**, 635–669 (2013). <https://doi.org/10.1613/jair.3870>

- [8] Ceylan, İ.İ., Lukasiewicz, T., Malizia, E., Molinaro, C., Vaicenavicius, A.: Explanations for negative query answers under existential rules. In: Calvanese, D., Erdem, E., Thielscher, M. (eds.) Proceedings of KR 2020. pp. 223–232. AAAI Press (2020). <https://doi.org/10.24963/kr.2020/23>
- [9] Del-Pinto, W., Schmidt, R.A.: ABox abduction via forgetting in \mathcal{ALC} . In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019. pp. 2768–2775. AAAI Press (2019). <https://doi.org/10.1609/aaai.v33i01.33012768>
- [10] Du, J., Qi, G., Shen, Y., Pan, J.Z.: Towards practical ABox abduction in large description logic ontologies. *Int. J. Semantic Web Inf. Syst.* **8**(2), 1–33 (2012). <https://doi.org/10.4018/jswis.2012040101>
- [11] Du, J., Wan, H., Ma, H.: Practical TBox abduction based on justification patterns. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 1100–1106 (2017), <http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14402>
- [12] Du, J., Wan, H., Ma, H.: Practical TBox abduction based on justification patterns. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 1100–1106 (2017), <http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14402>
- [13] Du, J., Wang, K., Shen, Y.: A tractable approach to ABox abduction over description logic ontologies. In: Brodley, C.E., Stone, P. (eds.) Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. pp. 1034–1040. AAAI Press (2014), <http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8191>
- [14] Echenim, M., Peltier, N., Sellami, Y.: A generic framework for implicate generation modulo theories. In: Galmiche, D., Schulz, S., Sebastiani, R. (eds.) Automated Reasoning - 9th International Joint Conference, IJCAR 2018, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, UK, July 14–17, 2018, Proceedings. Lecture Notes in Computer Science, vol. 10900, pp. 279–294. Springer (2018). https://doi.org/10.1007/978-3-319-94205-6_19, https://doi.org/10.1007/978-3-319-94205-6_19
- [15] Elsenbroich, C., Kutz, O., Sattler, U.: A case for abductive reasoning over ontologies. In: Proceedings of the OWLED’06 Workshop on OWL: Experiences and Directions (2006), http://ceur-ws.org/Vol-216/submission_25.pdf
- [16] Grau, B.C., Horrocks, I., Kazakov, Y., Sattler, U.: Modular reuse of ontologies: Theory and practice. *J. Artif. Intell. Res.* **31**, 273–318 (2008). <https://doi.org/10.1613/jair.2375>, <https://doi.org/10.1613/jair.2375>
- [17] Haifani, F., Koopmann, P., Tournet, S., Weidenbach, C.: Connection-minimal abduction in \mathcal{EL} via translation to FOL – technical report (2022). <https://doi.org/10.48550/ARXIV.2205.08449>, <https://arxiv.org/abs/2205.08449>
- [18] Haifani, F., Koopmann, P., Tournet, S., Weidenbach, C.: Experiment data for the paper ”Connection-minimal Abduction in EL via translation to FOL” (May 2022). <https://doi.org/10.5281/zenodo.6563656>, <https://doi.org/10.5281/zenodo.6563656>
- [19] Haifani, F., Tournet, S., Weidenbach, C.: Generalized completeness for SOS resolution and its application to a new notion of relevance. In: Platzer, A., Sutcliffe, G. (eds.) Automated Deduction - CADE 28 - 28th International Conference on Automated Deduction, Virtual Event, July 12–15, 2021, Proceedings. Lecture Notes in Computer Science, vol. 12699, pp. 327–343. Springer (2021). https://doi.org/10.1007/978-3-030-79876-5_19, https://doi.org/10.1007/978-3-030-79876-5_19
- [20] Halland, K., Britz, K.: ABox abduction in \mathcal{ALC} using a DL tableau. In: 2012 South African Institute of Computer Scientists and Information Technologists Conference, SAICSIT ’12. pp. 51–58 (2012). <https://doi.org/10.1145/2389836.2389843>

- [21] Horridge, M., Bechhofer, S.: The OWL API: a java API for OWL ontologies. *Semantic Web* **2**(1), 11–21 (2011). <https://doi.org/10.3233/SW-2011-0025>, <https://doi.org/10.3233/SW-2011-0025>
- [22] Horridge, M., Parsia, B., Sattler, U.: Explanation of OWL entailments in protege 4. In: Bizer, C., Joshi, A. (eds.) *Proceedings of the Poster and Demonstration Session at the 7th International Semantic Web Conference (ISWC2008)*, Karlsruhe, Germany, October 28, 2008. CEUR Workshop Proceedings, vol. 401. CEUR-WS.org (2008), http://ceur-ws.org/Vol-401/iswc2008pd_submission_47.pdf
- [23] Kazakov, Y., Klinov, P., Stupnikov, A.: Towards reusable explanation services in protege. In: Artale, A., Glimm, B., Kontchakov, R. (eds.) *Proceedings of the 30th International Workshop on Description Logics*, Montpellier, France, July 18–21, 2017. CEUR Workshop Proceedings, vol. 1879. CEUR-WS.org (2017), <http://ceur-ws.org/Vol-1879/paper31.pdf>
- [24] Kazakov, Y., Krötzsch, M., Simancik, F.: The incredible ELK - from polynomial procedures to efficient reasoning with \mathcal{EL} ontologies. *J. Autom. Reason.* **53**(1), 1–61 (2014). <https://doi.org/10.1007/s10817-013-9296-3>, <https://doi.org/10.1007/s10817-013-9296-3>
- [25] Klarman, S., Endriss, U., Schlobach, S.: ABox abduction in the description logic \mathcal{ALC} . *Journal of Automated Reasoning* **46**(1), 43–80 (2011). <https://doi.org/10.1007/s10817-010-9168-z>
- [26] Koopmann, P.: Signature-based abduction with fresh individuals and complex concepts for description logics. In: Zhou, Z. (ed.) *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19–27 August 2021*. pp. 1929–1935. *ijcai.org* (2021). <https://doi.org/10.24963/ijcai.2021/266>
- [27] Koopmann, P., Del-Pinto, W., Touret, S., Schmidt, R.A.: Signature-based abduction for expressive description logics. In: Calvanese, D., Erdem, E., Thielscher, M. (eds.) *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, KR 2020*. pp. 592–602. AAAI Press (2020). <https://doi.org/10.24963/kr.2020/59>
- [28] Matentzoglou, N., Parsia, B.: Bioportal snapshot 30.03.2017 (Mar 2017). <https://doi.org/10.5281/zenodo.439510>, <https://doi.org/10.5281/zenodo.439510>
- [29] Nabeshima, H., Iwanuma, K., Inoue, K., Ray, O.: SOLAR: an automated deduction system for consequence finding. *AI Commun.* **23**(2-3), 183–203 (2010). <https://doi.org/10.3233/AIC-2010-0465>, <https://doi.org/10.3233/AIC-2010-0465>
- [30] Parsia, B., Matentzoglou, N., Gonçalves, R.S., Glimm, B., Steigmiller, A.: The OWL reasoner evaluation (ORE) 2015 competition report. *J. Autom. Reason.* **59**(4), 455–482 (2017). <https://doi.org/10.1007/s10817-017-9406-8>
- [31] Pukancová, J., Homola, M.: Tableau-based ABox abduction for the \mathcal{ALCHO} description logic. In: *Proceedings of the 30th International Workshop on Description Logics* (2017), <http://ceur-ws.org/Vol-1879/paper11.pdf>
- [32] Pukancová, J., Homola, M.: The AAA ABox abduction solver. *Künstliche Intell.* **34**(4), 517–522 (2020). <https://doi.org/10.1007/s13218-020-00685-4>
- [33] Schlobach, S., Cornet, R.: Non-standard reasoning services for the debugging of description logic terminologies. In: Gottlob, G., Walsh, T. (eds.) *Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI 2003)*. pp. 355–362. Morgan Kaufmann, Acapulco, Mexico (2003), <http://ijcai.org/Proceedings/03/Papers/053.pdf>
- [34] Wei-Kleiner, F., Dragisic, Z., Lambrix, P.: Abduction framework for repairing incomplete \mathcal{EL} ontologies: Complexity results and algorithms. In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. pp. 1120–1127.

- AAAI Press (2014), <http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8239>
- [35] Weidenbach, C., Schmidt, R.A., Hillenbrand, T., Rusev, R., Topic, D.: System description: SPASS Version 3.0. In: Pfenning, F. (ed.) Automated Deduction - CADE-21, 21st International Conference on Automated Deduction, Bremen, Germany, July 17-20, 2007, Proceedings. Lecture Notes in Computer Science, vol. 4603, pp. 514–520. Springer (2007). https://doi.org/10.1007/978-3-540-73595-3_38, https://doi.org/10.1007/978-3-540-73595-3_38
- [36] Wos, L., Robinson, G., Carson, D.: Efficiency and completeness of the set of support strategy in theorem proving. *Journal of the ACM* **12**(4), 536–541 (1965)