On a Notion of Relevance

Fajar Haifani[1,2], Patrick Koopmann[3], Sophie Tourret[1], and Christoph
Weidenbach[1]

[1] Max Planck Institute for Informatics, Saarbrücken Germany
[2] Graduate School of Computer Science, Saarbrücken, Germany
[3] University of Dresden, Dresden, Germany

Abstract. We define a notion of relevance of a clause for proving a par-
ticular entailment by the resolution calculus. We think that our notion of
relevance is useful for explaining why an entailment holds. A clause is *rel-
evant* if there is no proof of the entailment without it. It is *semi-relevant*
if there is a proof of the entailment using it. It is *irrelevant* if it is not
needed in any proof. By using well-known translations of description log-
ics to first-order clause logic, we show that all three notions of relevance
are decidable for a number of description logics, including $\mathcal{EL}$ and $\mathcal{ALC}$.
We provide effective tests for (semi-)relevance. The *(semi-)relevance* of a
DL axiom is defined with respect to the (semi-)relevance of the respective
clauses resulting from the translation.

1 Introduction

The motivation for this work comes in particular from complexity management
in the car industry [22,5]. Here, the set of all buildable cars is described by logical
formulas. The formulas encode the available car parts and the rules how to put
these parts together on a certain level of abstraction. The different models of the
conjunction of the overall set of formulas correspond to all individual buildable
cars. A build of a concrete car is performed by proving its entailment with respect
to its specification. For example, we may prove that a car with a specific engine,
body trim, and suspension can be built. An important question that arises in
this context is the overall importance of a rule or a part in a specific car or the
space of all cars. This means given some rule or part, we want to know whether
there is a car where the part, rule is used for its build. Obviously, if this is not
the case, then the part, rule is not used anymore and can be removed from the
overall set of logical formulas. The car portfolio permanently changes. Therefore,
an inherent part of the engineering is to detect superfluous parts or rules. On
the formula level, this question is reflected by our notion of semi-relevance, i.e.,
is there a proof of an entailment (e.g., a concrete built of a car) using a specific
formula (part, rule). Our notion of relevance captures all parts, rules that are
mandatory for the existence of a car. This information is very useful for a product
car engineer to understand the overall car portfolio and its dependencies.

The paper is organized as follows: in Section 2 we introduce first-order logic, clausal resolution and translation techniques for description logics to first-order logic up to the expressivity of $\mathcal{ALC}$ [9]. Then, in Section 3, we introduce our notion of relevance on the basis of first-order clauses. With respect to a DL ontology relevance of an axiom is defined with respect to relevance of the clauses out of the translation. Section 4 develops effective tests for our notions of relevance. Basically, relevance can be decided by resolution and semi-relevance by resolution with a set-of-support strategy. For any logic enjoying the finite model property for which an a priori bound on the size of models can be computed, the test for (semi-)relevance is decidable. Section 5 is devoted to a detailed comparison between our notion of relevance and already existing notions in the DL context. The paper ends with a discussion on directions of further work, in Section 6.

2 Preliminaries

We consider a *first-order* language over a set of variables $\mathcal{V}$ and a signature $\Sigma = (\Omega, \Pi)$ where Ω is an infinite set of function symbols and Π is a finite non-empty set of predicate symbols. Terms, atoms, literals, clauses and formulas are defined in the usual way. We write x, y for variables, a, b for constants, P, Q, R for predicate symbols, t, s for terms, A, B for atoms, L, K for literals and C, D for clauses, all possibly indexed and with or without arguments depending on the context. Given a literal L, $\mathrm{comp}(L) = \neg A$ if $L = A$ and $\mathrm{comp}(L) = A$ if $L = \neg A$. As usual, a term not containing any variable is called a *ground term*.

Substitutions σ, λ map variables to terms and are the identity on all but finitely many variables. They are homomorphically extended to terms, literals, clauses, and sets, sequences thereof. Two terms s, t are unifiable if there is a substitution σ with $s\sigma = t\sigma$. A most general substitution with respect to instantiation is called a most general unifier, or mgu. The notion of unification is extended to atoms and literals as usual.

The resolution calculus consists of the two inference rules resolution and factoring operating on two sets of clauses N and S:

Resolution $(N, S) \Rightarrow_{RES} (N, S \cup \{(D \vee C)\sigma\})$
        if the clauses $D \vee A$ and $\neg B \vee C$ are either both in S or one is in S
        and one is in N and if $\sigma = \mathrm{mgu}(A, B)$

Factoring $(N, S \uplus \{C \vee L \vee K\}) \Rightarrow_{RES} (N, S \cup \{C \vee L \vee K\} \cup \{(C \vee L)\sigma\})$
        if $\sigma = \mathrm{mgu}(L, K)$

The clauses $D \vee A$ and $\neg B \vee C$ are the *parents* of $(D \vee C)\sigma$ and the clause $C \vee L \vee K$ is the *parent* of $(C \vee L)\sigma$.

Given a set of clauses N, the start state $(\emptyset, N)$ initializes the above rules for the application of the classical resolution calculus. Given a set of clauses $N = N_1 \uplus N_2$, the start state (N_1, N_2) initializes the above rules for the application of the set-of-support (SOS) strategy with set-of-support N_2.

Theorem 1 (Soundness and Completeness of Resolution [19,23]). *The resolution calculus is sound and refutationally complete. The SOS strategy for an unsatisfiable clause set $N = N_1 \uplus N_2$ is refutationally complete if N_1 is satisfiable and N_2 is the SOS.*

A number of description logics can be expressed in first-order logic via a suitable translation. Furthermore, there is a close relationship between certain modal logics and description logics. Table 1, following [9], shows the translation of $\mathcal{ALC}$ to first-order logic where in [9] this was shown for a modal logic that can express $\mathcal{ALC}$. We leave out the definitional form from [9] as it is not needed for the examples presented here. A translation of a DL ontology is the conjunction of all translated axioms in the TBox and ABox. Also, we use τ similarly to translate concepts, roles, TBox, ABox, and the whole ontology. Note that, our relevance notion is defined on the clausified axioms. For simplicity, we do not show its clausification in Table 1. The later DL examples have a translation which is almost in clause form.

Table 1: $\mathcal{ALC}$ First-Order Translation

Concepts	
$\tau(\top, x) = \top$	$\tau(\bot, x) = \bot$
$\tau(A, x) = A(x)$	$\tau(\neg C, x) = \neg\tau(C, x)$
$\tau(C \sqcap D, x) = \tau(C, x) \wedge \tau(D, x)$	$\tau(C \sqcup D, x) = \tau(C, x) \vee \tau(D, x)$
$\tau(\forall r.C, x) = \forall y.(\tau(r, x, y)) \rightarrow \tau(C, x))$	$\tau(\exists r.C, x) = \exists y.(\tau(r, x, y) \wedge \tau(C, x))$
Roles	
$\tau(r, x, y) = r(x, y)$	
TBox Axioms	
$\tau(C \sqsubseteq D) = \forall x.(\pi(C, x) \rightarrow \tau(D, x))$	
$\tau(C \equiv D) = \forall x.(\tau(C, x) \leftrightarrow \tau(D, x))$	
ABox Axioms	
$\tau(C(d)) = \tau(C, d)$	
$\tau(r(d_1, d_2)) = r(d_1, d_2)$	

3 A Notion of Relevance

A *derivation* of a clause C from a clause set N is a finite sequence $\pi = [C_1, \ldots, C_n]$ with $C_n = C$ such that for each $C_i \in \pi$ either: (i) $C_i \in N$ or (ii) there is a clause $C_j \in \pi$, $j < i$ such that C_i is the result of a Factoring inference from C_j or (iii) there are clauses $C_j, C_k \in \pi$, $j < k < i$ such that C_i is the result of a Resolution inference from C_j and C_k. The derivation π is called *connected* if in addition every clause is in π is connected, where connectedness in π is defined

by $C_n = C$ is connected, and a clause $C_i \in \pi$ is connected if it is the parent of a connected clause. A connected derivation of $\bot$ out of N is a *refutation (proof)* of N.

A derivation $\pi = [C_1, \ldots, C_n]$ is an SOS derivation from (N, S) if $(N, S) \Rightarrow^*_{RES} (N, S')$ and $C_i \in (N \cup S')$ for all i.

Definition 2 (Relevance). *Given an unsatisfiable set of clauses N, a clause $C \in N$ is* relevant *if for all refutations π of N it holds that $C \in \pi$. A clause $C \in N$ is* semi-relevant *if there exists a refutation π of N in which $C \in \pi$. A clause $C \in N$ is* irrelevant *if there is no refutation π of N in which $C \in \pi$.*

A different notion of relevance was previously defined in the context of propositional abduction [4]. The authors provide algorithms and complexity results for various abduction settings in the propositional logic context. In addition to the fact that our notion of relevance is defined with respect to first-order clauses, in their context of propositional abduction, if a propositional variable is relevant, it must be satisfiability preserving when added to the theory (clause set). In our case, if a clause $C \in N$ is (semi-)relevant, then N is unsatisfiable and $N \setminus \{C\}$ may be unsatisfiable as well.

With respect to the first-order translation of DL ontologies, see Table 1, the above definition on clauses translates as follows. Whenever we want to prove the entailment of a DL axiom ϕ from an ontology $\mathcal{O}$, $\mathcal{O} \models \phi$, we refute $\tau(\mathcal{O}) \cup \tau(\neg \phi)$ via resolution. A DL axiom is then *relevant* in deriving some entailment, if any refutation of the negation of the entailment after the respective translations contains at least one clause out of the translation of the axiom. A DL axiom is *semi-relevant* in deriving some entailment, if there is a refutation of the negation of the entailment after the respective translation that contains one clause out of the translation of the axiom. It is *irrelevant* if there is no such refutation that contains a clause out of the axiom. Note that a relevant DL axiom may translate into several clauses consisting of more than one semi-relevant clauses which are not relevant individually.

It may happen that different DL axioms produce the very same clause. In this case we silently assume a labelling of clauses, where all versions of the same clause but different labels are kept [13].

Example 3. As an illustration for our notion of relevance in the DL context, we consider an $\mathcal{ALC}$ ontology $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$, where

$$\mathcal{T} = \{\mathsf{LuxurySedan} \sqcap \exists \mathsf{hasEngine}.\mathsf{HighPerformanceEngine} \sqsubseteq \mathsf{PerformanceCar},$$
$$\mathsf{LuxurySedan} \sqcup \mathsf{PerformanceCar} \sqsubseteq \mathsf{ExecutiveCar}\}$$

The first concept inclusion says, a luxurious sedan with a high performance engine is a performance car. The second one expresses that luxurious sedans and performance cars are executive cars. In addition, we consider the following

ABox:

$$\mathcal{A} = \{ \ \mathsf{LuxurySedan(mercedes)},$$
$$\mathsf{hasEngine(mercedes, v8)},$$
$$\mathsf{HighPerformanceEngine(v8)}$$
$$\mathsf{PerformanceCar(lamborghini)}\}.$$

The first-order translation of $\mathcal{T}$ according to Table 1 results in

$$\neg\mathsf{LuxurySedan}(x) \vee \neg\mathsf{hasEngine}(x, z) \vee \neg\mathsf{HighPerformanceEngine}(z) \vee \mathsf{PerformanceCar}(x)$$
$$(\neg\mathsf{LuxurySedan}(x) \wedge \neg\mathsf{PerformanceCar}(x)) \vee \mathsf{ExecutiveCar}(x)$$

and we want to prove the entailment $\mathcal{O} \models \mathsf{ExecutiveCar(mercedes)}$. To find (semi-)relevant axioms for this entailment, we consider $\neg\mathsf{ExecutiveCar(mercedes)}$ and the translation of $\mathcal{O}$ to first-order logic. A refutation of the obtained formula is the following:

$$\pi_1 = [(1) : \neg\mathsf{LuxurySedan}(x) \vee$$
$$\neg\mathsf{hasEngine}(x, z) \vee \neg\mathsf{HighPerformanceEngine}(z) \vee$$
$$\mathsf{PerformanceCar}(x),$$
$$(2) : \neg\mathsf{PerformanceCar(x)} \vee \mathsf{ExecutiveCar(x)}$$
$$(3) : \mathsf{LuxurySedan(mercedes)},$$
$$(4) : \mathsf{hasEngine(mercedes, v8)}, (5) : \mathsf{HighPerformanceEngine(v8)},$$
$$(6) : \neg\mathsf{LuxurySedan}(x) \vee$$
$$\neg\mathsf{hasEngine}(x, z) \vee \neg\mathsf{HighPerformanceEngine}(z) \vee$$
$$\mathsf{ExecutiveCar}(x),$$
$$(7) : \neg\mathsf{hasEngine(mercedes}, z) \vee \neg\mathsf{HighPerformanceEngine}(z) \vee$$
$$\mathsf{ExecutiveCar(mercedes)},$$
$$(8) : \neg\mathsf{HighPerformanceEngine(v8)} \vee \mathsf{ExecutiveCar(mercedes)},$$
$$(9) : \mathsf{ExecutiveCar(mercedes)}$$
$$(10) : \neg\mathsf{ExecutiveCar(mercedes)}$$
$$(11) : \bot].$$

The clauses (1) to (5) are the initial clauses where we simply perform consecutive resolution steps between clause (1) and the other clauses (2) to (5) from $\tau(\mathcal{A})$. This will result in the respective clauses (6) to (9). Clause (9) is then refuted by the negated conjecture, (10), resulting in $\bot$. Additionally, π_2 is the only alternative proof for $\mathcal{O} \models \mathsf{ExecutiveCar(mercedes)}$ with different initial clauses.

$$\pi_2 = [(1) : \neg\mathsf{LuxurySedan(x)} \vee \mathsf{ExecutiveCar(x)}$$
$$(2) : \mathsf{LuxurySedan(mercedes)},$$
$$(3) : \mathsf{ExecutiveCar(mercedes)}$$
$$(4) : \neg\mathsf{ExecutiveCar(mercedes)}$$
$$(5) : \bot].$$

The concept inclusion LuxurySedan ⊔ PerformanceCar ⊑ ExecutiveCar is therefore relevant since both proofs must contain at least one clause out of its first-order translation. The ABox axiom LuxurySedan(mercedes) is also relevant. Moreover, the TBox axioms LuxurySedan ⊓ ∃hasEngine.(HighPerformanceEngine) ⊑ PerformanceCar and the ABox axioms LuxurySedan(mercedes), hasEngine(mercedes, v8), and HighPerformanceEngine(v8) are semi-relevant but not relevant. Finally, the ABox axiom PerformanceCar(lamborghini) is irrelevant.

4   Testing for Relevance

Our notion of relevance can be effectively tested for a number of description logics, including $\mathcal{EL}$ and $\mathcal{ALC}$.

Lemma 4 (Relevant Clauses). *Given an unsatisfiable set of clauses N, $C \in N$ is relevant if and only if $N \setminus \{C\}$ is satisfiable.*

Proof. If $N \setminus \{C\}$ is satisfiable then clearly C is contained in any refutation. On the over hand, if C is relevant then, by definition, there cannot be a refutation of N without C. Because of refutational completeness $N \setminus \{C\}$ must be satisfiable. ☐

The characterization of semi-relevant clauses that are not relevant is more complicated. In the sequel we will prove that given an unsatisfiable clause set N, a clause $C \in N$ is semi-relevant if there exists an SOS refutation with set-of-support set $\{C\}$, i.e., a derivation $(N \setminus \{C\}, \{C\}) \Rightarrow_{\text{RES}}^* (N', S' \cup \{\bot\})$. This result is not a consequence of Theorem 1 because if C is semi-relevant but not relevant then $N \setminus \{C\}$ is still unsatisfiable. Thus, there is no completeness guarantee for a refutation with set-of-support $\{C\}$ so far. The idea of our proof is to transform a (non-SOS) refutation using C into an SOS refutation with set-of-support $\{C\}$.

As an example, consider the unsatisfiable clause set:

$$N = \{ \ (1) : \neg B(b,a) \vee B(x_1, f(x_6)), \ (2) : \neg B(x_3, f(a)) \vee A(f(a)),$$
$$(3) : \neg A(x_4) \vee \neg B(b, x_4), \ (4) : B(b, x_2) \vee C(x_2), \ (5) : \neg C(x_5)\},$$

where the clause (1) is semi-relevant but not relevant. We attach unique numbers to clauses for reference and indicate resolution (factoring) steps by writing nRm (respectively nF) meaning that the clause is a result of a resolvent between clause n and clause m (respectively a factor out of clause n). A refutation including clause (1) is:

$$\pi_1 = [\ (2), \ (3), \ 2R3 \ (6) : \neg B(x_3, f(a)) \vee \neg B(b, f(a)), \ 6F \ (7) : \neg B(b, f(a)),$$
$$(1), \ (4), \ 1R4 \ (8) : B(x_1, f(x_6)) \vee C(a), \ 8R7 \ (9) : C(a), \ (5), \ 9R5 \ (10) : \bot],$$

where the inference $1R4 \ (8)$ is an SOS step with set-of-support $\{\neg B(b,a) \vee B(x_1, f(x_6))\}$ while the inference $2R3 \ (6)$ and $6F \ (7)$ are not SOS steps. The

refutation is depicted in the graph in Figure 1 where shaded clauses mark clauses contained in an respective SOS when started with SOS $\{(1)\}$.
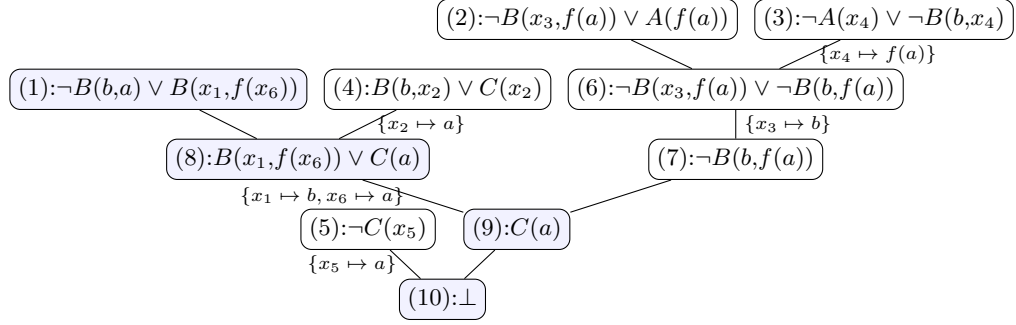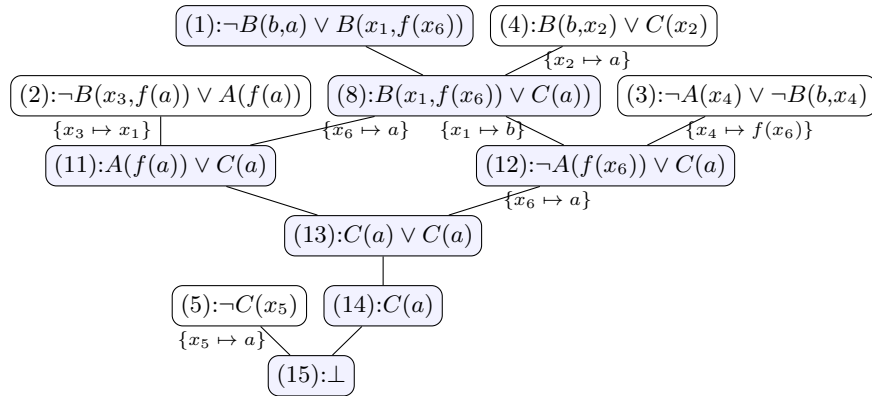

Fig. 1. Non-SOS Refutation with $\neg B(b, a) \vee B(x_1, f(x_6))$

Tracing back the literals in the original clause set resolved by $B(x_1, f(x_6))$ out of clause (8), the idea of our completeness proof is to start with these steps; for the example, resolving $B(x_1, f(x_6))$ in clause (1) with the appropriate literals from clauses (2) and (3). This then already leads to the following SOS refutation

$$\pi_2 = [\ (1),\ (4),\ 1R4(8) : B(x_1, f(x_6)) \vee C(a),\ (2),\ 2R8(11) : A(f(a)) \vee C(a),$$
$$3R8(12) : \neg A(f(x_6)) \vee C(a),\ 11R12(13) : C(a) \vee C(a),\ 13F(14) : C(a),$$
$$(5),\ 14R5(15) : \bot].$$

also depicted in the following graph, where shaded clauses mark clauses contained in the SOS.

Fig. 2. SOS refutation with $\neg B(b, a) \vee B(x_1, f(x_6))$

The transformation from the non-SOS refutation π_1 to the SOS refutation of π_2 is exactly the idea of the proof below of Theorem 5.

Clause (1) is not relevant, because there is also a refutation without it:

$$\pi_3 = [\,(4),\ (5),\ 4R5(16) : B(b, x_2),\ (2), 2R16(17) : A(f(a)),$$
$$(3),\ 3R17(18) : \neg B(b, f(a)),\ 16R18(19) : \bot\,]$$

Theorem 5 (Semi-Relevance Test). *Given an unsatisfiable clause set $N \uplus \{C\}$, the clause C is semi-relevant if and only if $(N, \{C\}) \Rightarrow^*_{RES} (N, S' \cup \{\bot\})$ via SOS.*

Proof. (Sketch) The backward implication holds by definition. To prove the forward implication let us assume the clause C is semi-relevant. We need to show that $(N, \{C\}) \Rightarrow^*_{\text{RES}} (N, S' \cup \{\bot\})$ via SOS. By definition, there is a resolution refutation π (without SOS) containing C. If π is also an SOS refutation we are done. If not, we subsequently transform π into an SOS proof via an inductive argument. For simplicity and without loss of generality, we assume that π has a tree structure, i.e., each clause is used exactly once in the refutation, and all clauses taken from $N \uplus \{C\}$ are variable disjoint. Therefore, there is a unique overall substitution σ such that $\pi\sigma$ is also a refutation where all mgus are identities. The underlying well-founded ordering for the inductive argument is the multiset of all distances (number of steps) from any clause $C \in \pi$ to $\bot$ where C is not generated by the SOS strategy. Now we pick a clause $nRm\,(l) : D$ in π, where $n, m < l$, l minimal, and m is contained in the SOS, but clause n is not in N and has not been generated by an SOS derivation. Let L be the literal of m resolved upon. Then we trace back comp(L) in π to the clauses out of N eventually generating the clause n. Now we resolve all these clauses first with variable-disjoint copies of m on comp(L). For each copy, σ is extended accordingly, i.e., the fresh extra variables are mapped to the same terms as their originals. Then we redo the derivation of n following exactly the steps out of π, resulting in π'. This derivation will then generate the clause D via SOS, possibly after some extra factoring steps. Thus π' gets strictly smaller with respect to the well-founded above defined measure. This transformation is possible, because any mgu occurring in π or in π' can always be instantiated to σ. Our transformation does not change the starting overall contradicting substitution σ. $\qquad\square$

Corollary 6 ((Semi)-Relevance Decidability). *If resolution is a decision procedure for some class of clause sets, then (semi-)relevance is decidable.*

Many description logics enjoy the *finite model property*, where the relevant finite model for some clause set can be explicitly a priori generated [9]. In this case the logics enjoy the *bounded model property*. In particular, if resolution is not a decision procedure for the logic under consideration, an explicit bound on the Herbrand universe is needed. In [9], the bounded model property of $\mathcal{EL}$, $\mathcal{ALC}$ and other logics is used to provide for a translation-based decision procedure for these logics. In general, this approach can however be used for many description logic that have the finite model property, including more expressive description logics such as $\mathcal{SHOI}$.

Lemma 7 ((Semi)-Relevance Decidability in Description Logics). *For ontologies in a DL that enjoys the bounded model property, (semi-)relevance of an axiom for a given property is decidable.*

Proof. We translate both the (negated) property and the ontology to first-order logic. For relevance, we first check satisfiability of the resulting clause set without the clauses from the axiom via resolution using the bounded model property. This terminates because we only have to consider terms out of the finite Herbrand base generated by these logics [9]. If the clause set is unsatisfiable, the axiom is either irrelevant or semi-relevant but not relevant. If this clause set is satisfiable the axiom is either relevant or irrelevant because then semi-relevance implies relevance. In both cases testing for semi-relevance provides the final classification of the axiom.

To test for semi-relevance, we perform an SOS resolution proof attempt where the set of support contains the clauses corresponding to the axiom. If this results in a refutation, the axiom is semi-relevant or relevant for the property depending on the previous test, otherwise it is irrelevant.

All resolution proof attempts terminate, because they can be stopped once the generated terms exceed the expected bounded Herbrand universe. $\qquad\square$

Example 8. We illustrate how the semi-relevant clauses may be lifted to semi-relevance on TBox axioms. Consider the ontology $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$ where:

$$\mathcal{T} = \{A \sqsubseteq B \sqcap F, (B \sqcap C) \sqcup G \sqsubseteq D, B \sqcup D \sqsubseteq E\},$$
$$\mathcal{A} = \{A(a), C(a)\},$$

and consider the entailment $\mathcal{O} \models E(a)$. The TBox $\mathcal{T}$ is translated into:

$$\tau(A \sqsubseteq B \sqcap F) = \{\neg A(x) \lor B(x), \neg A(x) \lor F(x)\},$$
$$\tau((B \sqcap C) \sqcup G \sqsubseteq D) = \{\neg B(x) \lor \neg C(x) \lor D(x), \neg G(x) \lor D(x)\},$$
$$\tau(B \sqcup D \sqsubseteq E) = \{\neg B(x) \lor E(x), \neg D(x) \lor E(x)\}.$$

For the above clauses, the bounded Herbrand universe consists of a single element: $\{a\}$. Either this clause set has a model with domain $\{a\}$ or it does not have a model at all. Furthermore, for this particular clause set resolution is already a decision procedure. Then the axioms $A \sqsubseteq B \sqcap F$ and $B \sqcup D \sqsubseteq E$ are relevant because any refutation of $E(a)$ will always contain at least one clauses from $\tau(A \sqsubseteq B \sqcap F)$ and $\tau(B \sqcup D \sqsubseteq E)$ respectively. Moreover $(B \sqcap C) \sqcup G \sqsubseteq D$ is semi-relevant because we can find a refutation where either $\neg B(x) \lor \neg C(x) \lor D(x)$ or $\neg G(x) \lor D(x)$ do not exist.

Note that here, relevance in clausal form differs when it is lifted to DL axiom. The axiom $B \sqcup D \sqsubseteq E$ is relevant but its translation consists of only semi-relevant (but not relevant) clauses.

5 Relations to Other Notions in Description Logics and Beyond

The notions *relevant axioms* and *semi-relevant axioms* are related to two problems in description logics: justifications and repairs. Given an ontology $\mathcal{O}$ and an axiom α s.t. $\mathcal{O} \models \alpha$, a *justification for* $\mathcal{O} \models \alpha$ is a subset-minimal set $\mathcal{J} \subseteq \mathcal{O}$ s.t. $\mathcal{J} \models \alpha$ [10,20]. A generalisation of justifications is the concept of a *MinA* [1], which denotes a minimal axiom set preserving some given property. The process of computing justifications and MinAs is known as *axiom pinpointing*. Justifications are classically used for explaining entailments observed by a reasoner, for which they are used for example in the ontology editor Protégé [14]. Axioms that occur in at least one justification have also been studied under the name *MinA-relevant axioms* [1,17]. In addition to MinA-relevancy, they even study a corresponding counting problem #MINA-RELEVANCE which is to count the number of MinAs in which a given axiom occurs. MinA-relevant axioms correspond to the case where this count is 1 or more, and MinA-irrelevant axioms are those whose count is 0, but the count can give a more detailed account on the relevance of axioms. If the count correspond to the number of MinAs for the entailment (the corresponding counting problem #MINA is also studied in [17]), then it occurs in every justification.

If an axiom occurs in a justification for α, in the translation into first order logic, the axiom will be involved in some proof for $\mathcal{O} \models \alpha$. Thus, every axiom in a justification is semi-relevant for this entailment. One may be tempted to identify MinA-relevant axioms with semi-relevant ones. Though in most cases, these notions will coincide, it is in fact possible for an axiom to be semi-relevant even if it does not occur in any justification, as the following example shows.

Example 9. Consider the ontology $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$ consisting of the TBox:

$$\mathcal{T} = \{A \sqsubseteq B, B \sqcap C \sqsubseteq D, B \sqcup D \sqsubseteq E\}$$

and the ABox $\mathcal{A} = \{A(a), C(a)\}$; and consider the entailment $\mathcal{O} \models E(a)$. There is only one justification for this entailment, namely $\{A \sqsubseteq B, B \sqcup D \sqsubseteq E, A(a)\}$, so that only those axioms are MinA-relevant for $E(a)$. In contrast, all axioms in the ontology are semi-relevant for $E(a)$. A first proof in first-order logic uses, in addition to the negation of $E(a)$ the axioms $A(a)$, $A \sqsubseteq B$ and $B \sqcup D \sqsubseteq E$:

$$[A(a), \neg A(x) \vee B(x), B(a), \neg B(x) \vee E(x), E(a), \neg E(a), \bot].$$

A second proof also uses $A(a)$, $A \sqsubseteq B$ and $B \sqcup D \sqsubseteq E$, but additionally includes $C(a)$ and $B \sqcap C \sqsubseteq D$:

$$[A(a), \neg A(x) \vee B(x), B(a), \neg B(x) \vee \neg C(x) \vee D(x), \neg C(a) \vee D(a), C(a),$$
$$D(a), \neg D(x) \vee E(x), E(a), \neg E(a), \bot].$$

As the example illustrates, our notion of relevance captures connections between axioms and entailments that are not captured by the classical notion of justification or MinA-relevance.

The notion of relevant axioms is also related to what has been studied in the field of propositional satisfiability under the name of *lean kernels* [11,12]: given an unsatisfiable set F of propositional clauses, the lean kernel $N_a(F) \subseteq N$ consists exactly of those clauses that are involved in at least one refutation proof of N in the resolution calculus, and thus, in our terminology, the set of semi-relevant clauses. This notion has been extended to description logics in [16], not only considering resolution as an inference procedure, but even arbitrary consequence-based reasoning procedures. [16] also discuss a generalisation of lean kernels called *MinA-preserving module*: given an ontology $\mathcal{O}$ and an axiom α, a MinA-preserving module for α in $\mathcal{O}$ is a subset $\mathcal{M} \subseteq \mathcal{O}$ s.t. every justification for $\mathcal{O} \models \alpha$ is a subset of $\mathcal{M}$. There is no minimality requirement, that is, a MinA-preserving module may contain axioms that do not occur in any justification. Because every axiom in a justification is also semi-relevant, the set of all semi-relevant axioms is also a MinA-preserving module. However, as we illustrated with the previous example, the set of semi-relevant axioms may capture more relations than a MinA-preserving module.

As semi-relevance is determined on the level of first-order clauses, there is also a connection to *laconic justifications* [7]. A laconic justification departs from a justification in that it does not have to be a subset of the original ontology, but may contain axioms that are the result of simplifying axioms in a justification, for instance by removing conjuncts and disjuncts that are irrelevant for the entailment. When considering the set of semi-relevant *clauses* instead of axioms, we obtain a similar level of precision as with laconic justifications.

Example 10. Consider the ontology:

$$\mathcal{O} = \{A \sqsubseteq B \sqcap C, D \sqcup B \sqsubseteq E\}.$$

The whole ontology is a justification for $\mathcal{O} \models A \sqsubseteq E$. However, the concepts C and D are *superfluous*, i.e., they are not needed for the entailment, so that the axioms would be simplified for the laconic justification. The laconic justification is thus $\mathcal{J} = \{A \sqsubseteq B, B \sqsubseteq E\}$.

Now consider the first-order translation of $\mathcal{O}$.

$$\neg A(x) \lor B(x), \neg A(x) \lor C(x), \neg D(x) \lor E(x), \neg B(x) \lor E(x) \tag{1}$$

Here, our notion would categorize $\{\neg A(x) \lor B(x), \neg B(x) \lor E(x)\}$ as *relevant*, while the clauses $\{\neg A(x) \lor C(x), \neg D(x) \lor E(x)\}$ would be *irrelevant*. The relevant clauses correspond to the axioms in the laconic justification. It remains to check whether this always holds or not, which we leave as future work.

While axioms in justifications correspond to semi-relevant axioms, an axiom that occurs in all justifications corresponds to a relevant axiom. When exhibiting justifications, it is helpful to understand whether they occur only in one justification or in several. This is for instance reflected in the visual presentation of justifications in Protégé [15], that indicates for each axiom in a justification in how many justifications it is included, or whether it is included in every justification.

Another important connection however is from relevant axioms to *repairs*. In ontology repair, the aim is to remove an unwanted entailment from an ontology. Specifically, a *repair* for an entailment $\mathcal{O} \models \alpha$ is a subset-minimal set $\mathfrak{R} \subseteq \mathcal{O}$ s.t. $\mathcal{O} \setminus \mathfrak{R} \not\models \alpha$ [8]. A repair for $\mathcal{O} \models \alpha$ always contains one axiom from every justification for $\mathcal{O} \models \alpha$, a fact that is used in the method for computing repairs based on Reiter's Minimal Hitting Set Algorithm [18]. It is not hard to see that a relevant axiom corresponds to the special case of a repair consisting of exactly one element. Rather than computing these repairs through the detour of justifications, our method provides a way of computing these repairs directly.

Two final formalisms worth mentioning are provenance semirings and axiom pinpointing formulas, as they both are often used to link inferences to entailments. Given an entailment α and an ontology $\mathcal{O}$, a *pinpointing formula* is a monotone propositional formula ϕ over the axioms in $\mathcal{O}$ (as propositional symbols) s.t. every model of the formula entails α, and the set of justifications corresponds exactly to the set of minimal models of ϕ. Thus, pinpointing formulas can be used to generate justifications, but also have other applications. Similar to our approach for computing semi-relevant axioms, there are methods for computing pinpointing formulas by intercepting reasoning procedures, as for instance done for $\mathcal{EL}$ in [2]. This idea of pinpointing formulas is very related to that of provenance semirings from the field of database access [6], recently investigated also in the context of DLs [3]. Here the idea is to compute a formula, called *provenance polynomial*, that links a query result or entailment to the database entries that are involved in computing the query.

6   Conclusion

We introduced the notions of relevance and semi-relevance in first-order logic, showed their effectiveness and discussed their relationship to existing notions known in the DL context. Our notion of relevance adds new aspects in understanding an ontology to the related DL notions of justifications and repairs. In particular, semi-relevant clauses appear to be particularly connected to laconic justifications and the computation of relevant axioms offers an alternative to existing repair methods.

The test procedure suggested in Section 4 relying on an explicit finite domain can be inefficient. It would be desirable to use the original first-order decision procedures, e.g., ordered resolution [21] for the (semi-)relevance tests. For relevance this can be obviously done. For semi-relevance this needs an extra argument, because the SOS strategy is not a priori compatible with ordered resolution. However, we believe that an extended concept of saturation can lead to a complete SOS procedure for ordered resolution. In any case, the design of an efficient semi-relevancy detection algorithm is the next considered step in this work.

References

1. Baader, F., Peñaloza, R.: Axiom pinpointing in general tableaux. J. Log. Comput. **20**(1), 5–34 (2010). https://doi.org/10.1093/logcom/exn058, `https://doi.org/10.1093/logcom/exn058`

2. Baader, F., Peñaloza, R., Suntisrivaraporn, B.: Pinpointing in the description logic EL^+. In: Hertzberg, J., Beetz, M., Englert, R. (eds.) KI 2007: Advances in Artificial Intelligence, 30th Annual German Conference on AI, KI 2007, Osnabrück, Germany, September 10-13, 2007, Proceedings. Lecture Notes in Computer Science, vol. 4667, pp. 52–67. Springer (2007). https://doi.org/10.1007/978-3-540-74565-5_7, `https://doi.org/10.1007/978-3-540-74565-5_7`

3. Bourgaux, C., Ozaki, A.: Querying attributed dl-lite ontologies using provenance semirings. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019. pp. 2719–2726. AAAI Press (2019). https://doi.org/10.1609/aaai.v33i01.33012719, `https://doi.org/10.1609/aaai.v33i01.33012719`

4. Eiter, T., Gottlob, G.: The complexity of logic-based abduction. Journal of the ACM **42**(1), 3–42 (1995)

5. Fetzer, C., Weidenbach, C., Wischnewski, P.: Compliance, functional safety and fault detection by formal methods. In: Margaria, T., Steffen, B. (eds.) Leveraging Applications of Formal Methods, Verification and Validation: Discussion, Dissemination, Applications - 7th International Symposium, ISoLA 2016, Imperial, Corfu, Greece, October 10-14, 2016, Proceedings, Part II. Lecture Notes in Computer Science, vol. 9953, pp. 626–632 (2016)

6. Green, T.J., Karvounarakis, G., Tannen, V.: Provenance semirings. In: Libkin, L. (ed.) Proceedings of the Twenty-Sixth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 11-13, 2007, Beijing, China. pp. 31–40. ACM (2007). https://doi.org/10.1145/1265530.1265535, `https://doi.org/10.1145/1265530.1265535`

7. Horridge, M., Parsia, B., Sattler, U.: Laconic and precise justifications in OWL. In: The Semantic Web - ISWC 2008, 7th International Semantic Web Conference, ISWC 2008, Karlsruhe, Germany, October 26-30, 2008. Proceedings. pp. 323–338 (2008). https://doi.org/10.1007/978-3-540-88564-1_21, `https://doi.org/10.1007/978-3-540-88564-1_21`

8. Horridge, M., Parsia, B., Sattler, U.: Explaining inconsistencies in OWL ontologies. In: Godo, L., Pugliese, A. (eds.) Scalable Uncertainty Management, Third International Conference, SUM 2009. Lecture Notes in Computer Science, vol. 5785, pp. 124–137. Springer (2009). https://doi.org/10.1007/978-3-642-04388-8_11, `https://doi.org/10.1007/978-3-642-04388-8_11`

9. Hustadt, U., Schmidt, R.A.: Using resolution for testing modal satisfiability and building models. Journal of Automed Reasoning **28**(2), 205–232 (2002)

10. Kalyanpur, A., Parsia, B., Horridge, M., Sirin, E.: Finding all justifications of OWL DL entailments. In: Aberer, K., Choi, K., Noy, N.F., Allemang, D., Lee, K., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. Lecture Notes in Computer Science, vol. 4825, pp. 267–280. Springer (2007). https://doi.org/10.1007/978-3-540-76298-0_20, `https://doi.org/10.1007/978-3-540-76298-0_20`

11. Kleine Büning, H., Kullmann, O.: Minimal unsatisfiability and autarkies. In: Biere, A., Heule, M., van Maaren, H., Walsh, T. (eds.) Handbook of Satisfiability, Frontiers in Artificial Intelligence and Applications, vol. 185, pp. 339–401. IOS Press (2009). https://doi.org/10.3233/978-1-58603-929-5-339, `https://doi.org/10.3233/978-1-58603-929-5-339`

12. Kullmann, O.: Investigations on autark assignments. Discret. Appl. Math. **107**(1-3), 99–137 (2000). https://doi.org/10.1016/S0166-218X(00)00262-6, `https://doi.org/10.1016/S0166-218X(00)00262-6`

13. Lev-Ami, T., Weidenbach, C., Reps, T.W., Sagiv, M.: Labelled clauses. In: Pfenning, F. (ed.) Automated Deduction - CADE-21, 21st International Conference on Automated Deduction, Bremen, Germany, July 17-20, 2007, Proceedings. LNCS, vol. 4603, pp. 311–327. Springer (2007)

14. Musen, M.A.: The protégé project: a look back and a look forward. AI Matters **1**(4), 4–12 (2015)

15. Musen, M.A.: The protégé project: a look back and a look forward. AI Matters **1**(4), 4–12 (2015). https://doi.org/10.1145/2757001.2757003, `https://doi.org/10.1145/2757001.2757003`

16. Peñaloza, R., Mencía, C., Ignatiev, A., Marques-Silva, J.: Lean kernels in description logics. In: Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler, P., Hartig, O. (eds.) The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I. Lecture Notes in Computer Science, vol. 10249, pp. 518–533 (2017). https://doi.org/10.1007/978-3-319-58068-5_32, `https://doi.org/10.1007/978-3-319-58068-5_32`

17. Peñaloza, R., Sertkaya, B.: Understanding the complexity of axiom pinpointing in lightweight description logics. Artif. Intell. **250**, 80–104 (2017). https://doi.org/10.1016/j.artint.2017.06.002, `https://doi.org/10.1016/j.artint.2017.06.002`

18. Reiter, R.: A theory of diagnosis from first principles. Artif. Intell. **32**(1), 57–95 (1987). https://doi.org/10.1016/0004-3702(87)90062-2, `https://doi.org/10.1016/0004-3702(87)90062-2`

19. Robinson, J.A.: A machine-oriented logic based on the resolution principle. J. ACM **12**(1), 23–41 (1965). https://doi.org/10.1145/321250.321253, `http://doi.acm.org/10.1145/321250.321253`

20. Schlobach, S., Cornet, R.: Non-standard reasoning services for the debugging of description logic terminologies. In: Gottlob, G., Walsh, T. (eds.) IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003. pp. 355–362. Morgan Kaufmann (2003), `http://ijcai.org/Proceedings/03/Papers/053.pdf`

21. Schmidt, R.A.: Decidability by resolution for propositional modal logics. J. Autom. Reasoning **22**(4), 379–396 (1999). https://doi.org/10.1023/A:1006043519663, `https://doi.org/10.1023/A:1006043519663`

22. Sinz, C., Kaiser, A., Küchlin, W.: Formal methods for the validation of automotive product configuration data. AI EDAM **17**(1), 75–97 (2003). https://doi.org/10.1017/S0890060403171065, `https://doi.org/10.1017/S0890060403171065`

23. Wos, L., Robinson, G.A., Carson, D.F.: Efficiency and completeness of the set of support strategy in theorem proving. J. ACM **12**(4), 536–541 (1965). https://doi.org/10.1145/321296.321302, `https://doi.org/10.1145/321296.321302`