Binary Absorption in Tableaux-Based Reasoning for Description Logics

Alexander K. Hudek and Grant Weddell
David R. Cheriton School of Computer Science
University of Waterloo
{akhudek,gweddell}@cs.uwaterloo.ca

1 Introduction

A fundamental problem in Description Logics (DLs) is *satisfiability*, the problem of checking if a given DL *terminology* $\mathcal{T}$ remains sufficiently unconstrained to enable at least one instance of a given DL *concept* C to exist. It has been known for some time that *lazy unfolding* is an important optimization technique in model building algorithms for satisfiability [2]. It is also imperative for large terminologies to be manipulated by an *absorption generation* process to maximize the benefits of lazy unfolding in such algorithms, thereby reducing the combinatorial effects of disjunction in underlying chase procedures [5]. In this paper, we propose a generalization of the absorption theory and algorithms developed by Horrocks and Tobies [6, 7]. The generalization, called *binary absorption*, makes it possible for lazy unfolding to be used for parts of terminologies not handled by current absorption algorithms and theory.

The basic idea of binary absorption is to avoid the need to *internalize* (at least some of the) terminological axioms of the form

$$(A_1 \sqcap A_2) \sqsubseteq C,$$

where the A_i denote *primitive concepts* and C a general concept. This idea, coupled with equivalences and another idea relating to "role absorptions" developed by Tsarkov and Horrocks [8], makes it possible for an algorithm to absorb, e.g., the definition

$$\begin{aligned}\text{SPECIALCLIENT} \doteq\ & \text{CLIENT} \sqcap \\ & (\exists Buy.(\text{EXPENSIVE} \sqcup \text{PROFITABLE})) \sqcap \\ & (\exists Recommend^{-}.\text{TRUSTEDCLIENT})\end{aligned}$$

as the set of axioms

$$
\begin{array}{rcl}
\text{SPECIALCLIENT} & \sqsubseteq & \text{CLIENT} \sqcap \\
& & (\exists Buy.(\text{EXPENSIVE} \sqcup \text{PROFITABLE})) \sqcap \\
& & (\exists Recommend^{-}.\text{TRUSTEDCLIENT}) \\
\text{EXPENSIVE} & \sqsubseteq & \text{A1} \\
\text{PROFITABLE} & \sqsubseteq & \text{A1} \\
\text{A1} & \sqsubseteq & \forall Buy^{-}.\text{A2} \\
\text{CLIENT} \sqcap \text{A2} & \sqsubseteq & \text{A3} \\
\text{TRUSTEDCLIENT} & \sqsubseteq & \forall Recommend.\text{A4} \\
\text{A3} \sqcap \text{A4} & \sqsubseteq & \text{SPECIALCLIENT}
\end{array}
$$

in which the primitive concepts A1, A2, A3 and A4 are new internal primitive concepts introduced by the absorption algorithm.

There is another reason that binary absorption proves useful, beyond the well-documented advantages of reducing the need for internalization of general terminological axioms. In particular, Palacios has explored the possibility of a tighter integration of conjunctive query optimization and DL consistency checking in which a query is mapped to a concept in such a way that view integration becomes an automatic consequence of an initial *chase phase* for optimization [4, 9]. When used for such purposes, it becomes crucial for DL reasoners to support special case (and likely incomplete) *deterministic* modes of model building. Binary absorption enables such modes to incorporate reasoning about *materialized views* in a terminology, such as SPECIALCLIENT above.

The organization of the paper is as follows. The next section is a review of the basic definitions introduced by Horrocks and Tobies [6] for the notion of a DL, of terminologies and satisfiability, and an important abstraction that relates to model building algorithms for satisfiability: the notion of a *witness*. In Section 3, we define binary absorptions, and give a related lemma that establishes an additional sufficiency condition for correct binary absorptions. (Note that existing sufficiency conditions given by Horrocks and Tobies [6, 7] for correct non-binary absorptions are inherited in our formulation.) In Section 4, we present an absorption algorithm for binary absorptions that derives from earlier procedures [1, 6, 7]. Our summary comments follow in Section 5.

2 Preliminaries

The definitions and lemmas in this section are largely reproduced from Horrocks and Tobies [7]. The main difference is in our formulation of a description logic immediately following in which we have an additional stipulation that the logic includes the concept forming operations of $\mathcal{ALCI}$.

Definition 2.1. (Description Logic) *Let* L *be a DL based on infinite sets of atomic concepts* NC *and atomic roles* NR. *We identify* L *with the set of its well-formed concepts and require* L *to be closed under sub-concepts and the concept forming operations of dialect $\mathcal{ALCI}$: we require that if* NR *contains R and* L *contains C_1 and C_2, then* L *also contains $\neg C_1$, $C_1 \sqcap C_2$, $C_1 \sqcup C_2$, $\top$, $\bot$, $\exists R.C_1$, $\exists R^{-}.C_1$, $\forall R.C_1$ and $\forall R^{-}.C_1$.*

An interpretation $\mathcal{I}$ is a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set and $\cdot^{\mathcal{I}}$ is a function mapping NC *to subsets of $\Delta^{\mathcal{I}}$ and* NR *to subsets of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. Each* L *is associated with a set* Int(L) *of* admissible *interpretations that (1) must be closed under isomorphisms, and that (2) satisfies $\mathcal{I} \in \mathrm{Int(L)} \Leftrightarrow \mathcal{I}' \in \mathrm{Int(L)}$ for any two interpretations $\mathcal{I}$ and $\mathcal{I}'$ that agree on* NR. *Each* L *must also come with a semantics that allows any $\mathcal{I} \in \mathrm{Int(L)}$ to be extended to each concept $C \in \mathrm{L}$ in a way that satisfies the following conditions:*

(I1) the concept forming operations of $\mathcal{ALCI}$ are mapped in the standard way, and

(I2) the interpretation $C^{\mathcal{I}}$ of a compound concept $C \in \mathrm{L}$ depends only on the interpretation of those atomic concepts and roles that appear syntactically in C.

Definition 2.2. (TBox, Satisfiability) *A* TBox *$\mathcal{T}$ for* L *is a finite set of axioms of the form $C_1 \sqsubseteq C_2$ or $C_1 \doteq C_2$ where $C_i \in \mathrm{L}$. If $\mathcal{T}$ contains $A \sqsubseteq C$ or $A \doteq C$ for some $A \in \mathrm{NC}$, then we say that A* is defined in *$\mathcal{T}$.*

Let L *be a DL and $\mathcal{T}$ a TBox. An interpretation $\mathcal{I} \in \mathrm{Int(L)}$ is a* model *of $\mathcal{T}$, written $\mathcal{I} \models \mathcal{T}$, iff $C_1^{\mathcal{I}} \subseteq C_2^{\mathcal{I}}$ holds for each $C_1 \sqsubseteq C_2 \in \mathcal{T}$, and $C_1^{\mathcal{I}} = C_2^{\mathcal{I}}$ holds for each $C_1 \doteq C_2 \in \mathcal{T}$. A concept $C \in \mathrm{L}$ is* satisfiable *with respect to a TBox $\mathcal{T}$ iff there is an $\mathcal{I} \in \mathrm{Int(L)}$ such that $\mathcal{I} \models \mathcal{T}$ and such that $C^{\mathcal{I}} \neq \emptyset$.*

A TBox $\mathcal{T}$ is called primitive *iff it consists entirely of axioms of the form $A \doteq C$ with $A \in \mathrm{NC}$, each $A \in \mathrm{NC}$ appears in at most one left hand side of an axiom, and $\mathcal{T}$ is* acyclic. *Acyclicity is defined as follows: $A_1 \in \mathrm{NC}$* directly uses *$A_2 \in \mathrm{NC}$ if $A_1 \doteq C \in \mathcal{T}$ and A_2 occurs in C;* uses *is the transitive closure of "directly uses". Then $\mathcal{T}$ is* acyclic *if there is no $A \in \mathrm{NC}$ that uses itself.*

Model building algorithms for checking the satisfaction of a concept C operate by manipulating an internal data structure (e.g., in the form of a node and edge labelled rooted tree with "back edges"). The data structure "encodes" a *partial description* of (eventual) interpretations $\mathcal{I}$ for which $C^{\mathcal{I}}$ will be non-empty. Such a partial description will almost always abstract details on class membership for hypothetical elements of $\Delta^{\mathcal{I}}$ and on details relating to the interpretation of roles. To talk formally about absorption and lazy evaluation, it is necessary to codify the idea of a partial description. Horrocks and Tobies have done this by introducing the following notion of a *witness*, of an interpretation that *stems* from a witness, and of what it means for a witness to be *admissible* with respect to a given terminology.

Definition 2.3. (Witness) *Let* L *be a DL and $C \in \mathrm{L}$ a concept. A* witness *$\mathcal{W} = (\Delta^{\mathcal{W}}, \cdot^{\mathcal{W}}, \mathcal{L}^{\mathcal{W}})$ for C consists of a non-empty set $\Delta^{\mathcal{W}}$, a function $\cdot^{\mathcal{W}}$ that maps* NR *to subsets of $\Delta^{\mathcal{W}} \times \Delta^{\mathcal{W}}$, and a function $\mathcal{L}^{\mathcal{W}}$ that maps $\Delta^{\mathcal{W}}$ to subsets of* L *such that:*

(W1) there is some $x \in \Delta^{\mathcal{W}}$ with $C \in \mathcal{L}^{\mathcal{W}}(x)$,

(W2) there is an interpretation $\mathcal{I} \in \mathrm{Int(L)}$ that stems *from $\mathcal{W}$, and*

(W3) for each interpretation $\mathcal{I} \in \mathrm{Int(L)}$ that stems *from $\mathcal{W}$, $x \in C^{\mathcal{I}}$ if $C \in \mathcal{L}^{\mathcal{W}}(x)$.*

An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *is said to* stem *from* $\mathcal{W}$ *if* $\Delta^{\mathcal{I}} = \Delta^{\mathcal{W}}$, $\cdot^{\mathcal{I}}|_{\mathrm{NR}} = \cdot^{\mathcal{W}}$, *and for each* $A \in \mathrm{NC}$, $A \in \mathcal{L}^{\mathcal{W}}(x)$ *implies* $x \in A^{\mathcal{I}}$ *and* $\neg A \in \mathcal{L}^{\mathcal{W}}(x)$ *implies* $x \notin A^{\mathcal{I}}$.

A witness $\mathcal{W}$ *is called* admissible *with respect to a TBox* $\mathcal{T}$ *if there is an interpretation* $\mathcal{I} \in \mathrm{Int}(\mathrm{L})$ *that stems from* $\mathcal{W}$ *with* $\mathcal{I} \models \mathcal{T}$.

The important properties satisfied by a witness are captured by the following lemmas by Horrocks and Tobies [7].

Lemma 2.1. *Let* L *be a DL. A concept* $C \in \mathrm{L}$ *is satisfiable w.r.t. a TBox* $\mathcal{T}$ *iff it has a witness that is admissible w.r.t.* $\mathcal{T}$.

Lemma 2.2. *Let* L, C, $\mathcal{T}$ *and* $\mathcal{W}$ *be a DL, a concept in* L, *a TBox for* L *and a witness for* C, *respectively. Then* $\mathcal{W}$ *is admissible w.r.t.* $\mathcal{T}$ *if, for each* $x \in \Delta^{\mathcal{W}}$:

$$\begin{array}{lll} C_1 \sqsubseteq C_2 \in \mathcal{T} & \textit{implies} & \neg C_1 \sqcup C_2 \in \mathcal{L}^{\mathcal{W}}(x), \\ C_1 \doteq C_2 \in \mathcal{T} & \textit{implies} & \neg C_1 \sqcup C_2 \in \mathcal{L}^{\mathcal{W}}(x) \textit{ and} \\ C_1 \doteq C_2 \in \mathcal{T} & \textit{implies} & C_1 \sqcup \neg C_2 \in \mathcal{L}^{\mathcal{W}}(x). \end{array}$$

3 Binary Absorptions

Our generalization of the notion of an *absorption* developed by Horrocks and Tobies [6, 7] is given as follows.

Definition 3.1. (Binary Absorption) *Let* L *and* $\mathcal{T}$ *be a DL and a TBox, respectively. A* binary absorption *of* $\mathcal{T}$ *is a pair of TBoxes* $(\mathcal{T}_u, \mathcal{T}_g)$ *such that* $\mathcal{T} \equiv \mathcal{T}_u \cup \mathcal{T}_g$ *and* $\mathcal{T}_u$ *contains only axioms of the form* $A_1 \sqsubseteq C$, *the form* $\neg A_1 \sqsubseteq C$ *or the form* $(A_1 \sqcap A_2) \sqsubseteq C$, *where* A_1 *and* A_2 *are primitive concepts occurring in* NC.

A binary absorption $(\mathcal{T}_u, \mathcal{T}_g)$ *of* $\mathcal{T}$ *is called* correct *if it satisfies the following condition. For each witness* $\mathcal{W}$ *and* $x \in \Delta^{\mathcal{W}}$, *if*

$$\begin{array}{rll} (A_1 \sqcap A_2) \sqsubseteq C \in \mathcal{T}_u \textit{ and } \{A_1, A_2\} \subseteq \mathcal{L}^{\mathcal{W}}(x) & \textit{implies} & C \in \mathcal{L}^{\mathcal{W}}(x), \\ A \sqsubseteq C \in \mathcal{T}_u \textit{ and } A \in \mathcal{L}^{\mathcal{W}}(x) & \textit{implies} & C \in \mathcal{L}^{\mathcal{W}}(x), \\ \neg A \sqsubseteq C \in \mathcal{T}_u \textit{ and } \neg A \in \mathcal{L}^{\mathcal{W}}(x) & \textit{implies} & C \in \mathcal{L}^{\mathcal{W}}(x), \\ C_1 \sqsubseteq C_2 \in \mathcal{T}_g & \textit{implies} & \neg C_1 \sqcup C_2 \in \mathcal{L}^{\mathcal{W}}(x), \\ C_1 \doteq C_2 \in \mathcal{T}_g & \textit{implies} & \neg C_1 \sqcup C_2 \in \mathcal{L}^{\mathcal{W}}(x) \textit{ and} \\ C_1 \doteq C_2 \in \mathcal{T}_g & \textit{implies} & C_1 \sqcup \neg C_2 \in \mathcal{L}^{\mathcal{W}}(x), \end{array} \tag{1}$$

then $\mathcal{W}$ *is admissible w.r.t.* $\mathcal{T}$. *A witness that satisfies* (1) *will be called* unfolded.

The distinguishing feature of binary absorption is the addition of the first implication in condition (1). This allows additional axioms in $\mathcal{T}_u$ to be dealt with in a deterministic manner, as we illustrate in our introductory example. If a label of a node contains neither primitive concept A_1 nor A_2, then *nothing* further is required. There is, however, a new requirement for computing the intersection of sets of primitive concepts during model building when a new primitive concept is added to a node. (For this problem, we refer the reader to algorithms developed by Demaine, Lopez-Ortiz and Munro [3].)

The next three lemmas by Horrocks and Tobies [7] hold without modification.

Lemma 3.1. *Let $(\mathcal{T}_u, \mathcal{T}_g)$ be a correct binary absorption of $\mathcal{T}$. For any $C \in \mathrm{L}$, C has a witness that is admissible w.r.t. $\mathcal{T}$ iff C has an unfolded witness.*

Lemma 3.2. *Let $\mathcal{T}$ be a primitive TBox and $\mathcal{T}_u$ defined as*

$$\{A \sqsubseteq C, \neg A \sqsubseteq \neg C \mid A \doteq C \in \mathcal{T}\}.$$

Then $(\mathcal{T}_u, \emptyset)$ is a correct absorption of $\mathcal{T}$.

Lemma 3.3. *Let $(\mathcal{T}_u, \mathcal{T}_g)$ be a correct absorption of a TBox $\mathcal{T}$.*

1. *If $\mathcal{T}'$ is an arbitrary TBox, then $(\mathcal{T}_u, \mathcal{T}_g \cup \mathcal{T}')$ is a correct absorption of $\mathcal{T} \cup \mathcal{T}'$.*

2. *If $\mathcal{T}'$ is a TBox that consists entirely of axioms of the form $A \sqsubseteq C$, where $A \in \mathrm{NC}$ and A is not defined in $\mathcal{T}_u$, then $(\mathcal{T}_u \cup \mathcal{T}', \mathcal{T}_g)$ is a correct absorption of $\mathcal{T} \cup \mathcal{T}'$.*

The main benefit of binary absorptions is that they allow the following additional sufficiency condition for correct absorptions.

Lemma 3.4. *Let $(\mathcal{T}_u, \mathcal{T}_g)$ be a correct absorption of a TBox $\mathcal{T}$. If $\mathcal{T}'$ is a TBox that consists entirely of axioms of the form $(A_1 \sqcap A_2) \sqsubseteq C$, where $\{A_1, A_2\} \subseteq \mathrm{NC}$ and where neither A_1 nor A_2 is defined in $\mathcal{T}_u$, then $(\mathcal{T}_u \cup \mathcal{T}', \mathcal{T}_g)$ is a correct absorption of $\mathcal{T} \cup \mathcal{T}'$.*

Proof. Observe that $\mathcal{T}_u \cup \mathcal{T}_g \cup \mathcal{T}' \equiv \mathcal{T} \cup \mathcal{T}'$ holds trivially. Let $C \in \mathrm{L}$ be a concept and $\mathcal{W}$ be an unfolded witness for C w.r.t. the absorption $(\mathcal{T}_u \cup \mathcal{T}', \mathcal{T}_g)$. From $\mathcal{W}$, define a new witness $\mathcal{W}'$ for C by setting $\Delta^{\mathcal{W}'} = \Delta^{\mathcal{W}}$, $\cdot^{\mathcal{W}'} = \cdot^{\mathcal{W}}$, and defining $\mathcal{L}^{\mathcal{W}'}$ to be the function that, for every $x \in \Delta^{\mathcal{W}'}$, maps x to the set

$$\begin{array}{rl} \mathcal{L}^{\mathcal{W}}(x) & \cup \ \{\neg A_1, \neg A_2 \mid (A_1 \sqcap A_2) \sqsubseteq C' \in \mathcal{T}', \{A_1, A_2\} \cap \mathcal{L}^{\mathcal{W}}(x) = \emptyset\} \\ & \cup \ \{\neg A_1 \mid (A_1 \sqcap A_2) \sqsubseteq C' \in \mathcal{T}', A_1 \notin \mathcal{L}^{\mathcal{W}}(x), A_2 \in \mathcal{L}^{\mathcal{W}}(x)\} \\ & \cup \ \{\neg A_2 \mid (A_1 \sqcap A_2) \sqsubseteq C' \in \mathcal{T}', A_1 \in \mathcal{L}^{\mathcal{W}}(x), A_2 \notin \mathcal{L}^{\mathcal{W}}(x)\}. \end{array}$$

It is easy to see that $\mathcal{W}'$ is also unfolded w.r.t. the absorption $(\mathcal{T}_u \cup \mathcal{T}', \mathcal{T}_g)$. This implies that $\mathcal{W}'$ is also unfolded w.r.t. the (smaller) absorption $(\mathcal{T}_u, \mathcal{T}_g)$. Since $(\mathcal{T}_u, \mathcal{T}_g)$ is a correct absorption of $\mathcal{T}$, there exists an interpretation $\mathcal{I}$ stemming from $\mathcal{W}'$ such that $\mathcal{I} \models \mathcal{T}$. We show that $\mathcal{I} \models \mathcal{T}'$ also holds. Assume $\mathcal{I} \not\models \mathcal{T}'$. Then there is an axiom $(A_1 \sqcap A_2) \sqsubseteq C_1 \in \mathcal{T}'$ and an $x \in \Delta^{\mathcal{I}}$ such that $x \in (A_1 \sqcap A_2)^{\mathcal{I}}$ but $x \notin C_1^{\mathcal{I}}$. By construction of $\mathcal{W}'$, $x \in (A_1 \sqcap A_2)^{\mathcal{I}}$ implies $\{A_1, A_2\} \subseteq \mathcal{L}^{\mathcal{W}'}(x)$ because otherwise $\{\neg A_1, \neg A_2\} \cap \mathcal{L}^{\mathcal{W}'}(x) \neq \emptyset$ would hold in contradiction to (W3). Then, since $\mathcal{W}'$ is unfolded, $C_1 \in \mathcal{L}^{\mathcal{W}'}(x)$, which, again, by (W3), implies $x \in C_1^{\mathcal{I}}$, a contradiction.

Hence, we have shown that there exists an interpretation $\mathcal{I}$ stemming from $\mathcal{W}'$ such that $\mathcal{I} \models \mathcal{T}_u \cup \mathcal{T}' \cup \mathcal{T}_g$. By construction of $\mathcal{W}'$, any interpretation stemming from $\mathcal{W}'$ also stems from $\mathcal{W}$, hence $\mathcal{W}$ is admissible w.r.t. $\mathcal{T} \cup \mathcal{T}'$. □

4 A Binary Absorption Algorithm

In this section, we present a two-phase algorithm for generating binary absorptions that derives from the absorption algorithm for the FaCT system outlined in earlier work [1, 6, 7]. Our algorithm also incorporates role absorption, similar to the work of Tsarkov and Horrocks [8]. The notable differences in our algorithm happen in the second phase during which

- an opportunity for unary absorption now has a very low priority,
- binary absorption replaces unary absorption in becoming a high priority, and
- primitive concepts are introduced by the procedure to enable further binary absorptions and to absorb existential role restrictions.

The algorithm is given a TBox $\mathcal{T}$ containing arbitrary axioms, and proceeds by constructing four TBoxes, $\mathcal{T}_g, \mathcal{T}_{prim}, \mathcal{T}_{uinc}$, and $\mathcal{T}_{binc}$, such that: $\mathcal{T} \equiv \mathcal{T}_g \cup \mathcal{T}_{prim} \cup \mathcal{T}_{uinc} \cup \mathcal{T}_{binc}$, $\mathcal{T}_{prim}$ is primitive, $\mathcal{T}_{uinc}$ consists only of axioms of the form $A_1 \sqsubseteq C$, and $\mathcal{T}_{binc}$ consists only of axioms of the form $(A_1 \sqcap A_2) \sqsubseteq C$, where $\{A_1, A_2\} \subseteq \text{NC}$ and neither A_1 nor A_2 are defined in $\mathcal{T}_{prim}$. Here, $\mathcal{T}_{uinc}$ contains unary absorptions and $\mathcal{T}_{binc}$ contains binary absorptions.

In the first phase, we move as many axioms as possible from $\mathcal{T}$ into $\mathcal{T}_{prim}$. We initialize $\mathcal{T}_{prim} = \emptyset$ and process each axiom $X \in \mathcal{T}$ as follows.

1. If X is of the form $A \doteq C$, A is not defined in $\mathcal{T}_{prim}$, and $\mathcal{T}_{prim} \cup \{X\}$ is primitive, then move X to $\mathcal{T}_{prim}$.
2. If X is of the form $A \doteq C$, then remove X from $\mathcal{T}$ and replace it with axioms $A \sqsubseteq C$ and $\neg A \sqsubseteq \neg C$.
3. Otherwise, leave X in $\mathcal{T}$.

In the second phase, we process axioms in $\mathcal{T}$, either by simplifying them or by placing absorbed components in either $\mathcal{T}_{uinc}$ or $\mathcal{T}_{binc}$. We place components that cannot be absorbed in $\mathcal{T}_g$. To ease axiom manipulation, we introduce a set representation. We let $\mathbf{G} = \{C_1, \ldots, C_n\}$ represent the axiom $\top \sqsubseteq (C_1 \sqcup \ldots \sqcup C_n)$. When removing an axiom from $\mathcal{T}$, we automatically convert it to a set $\mathbf{G}$. Similarly, when adding $\mathbf{G}$ to $\mathcal{T}$, we automatically convert it out of set notation. Details are as follows.

1. If $\mathcal{T}$ is empty, then return the binary absorption

   $$(\{A \sqsubseteq C, \neg A \sqsubseteq \neg C \mid A \doteq C \in \mathcal{T}_{prim}\} \cup \mathcal{T}_{uinc} \cup \mathcal{T}_{binc}, \mathcal{T}_g).$$

   Otherwise, remove an axiom $\mathbf{G}$ from $\mathcal{T}$.
2. Try to simplify $\mathbf{G}$.

   (a) If there is some $\neg C \in \mathbf{G}$ such that C is not a primitive concept, then add $(\mathbf{G} \cup \text{NNF}(\neg C) \setminus \{\neg C\}$ to $\mathcal{T}$, where the function $\text{NNF}(\cdot)$ converts concepts to negation normal form. Return to Step 1.

(b) If there is some $C \in \mathbf{G}$ such that C is of the form $(C_1 \sqcap C_2)$, then add both $(\mathbf{G} \cup \{C_1\}) \setminus \{C\}$ and $(\mathbf{G} \cup \{C_2\}) \setminus \{C\}$ to $\mathcal{T}$. Return to Step 1.

(c) If there is some $C \in \mathbf{G}$ such that C is of the form $C_1 \sqcup C_2$, then apply associativity by adding $(\mathbf{G} \cup \{C_1, C_2\}) \setminus \{C_1 \sqcup C_2\}$ to $\mathcal{T}$. Return to Step 1.

3. Try to partially absorb $\mathbf{G}$.

   (a) If $\{\neg A_1, \neg A_2\} \subset \mathbf{G}$, $(A_3 \sqcap A_4) \sqsubseteq A_1 \in \mathcal{T}_{binc}$, and A_2 is not defined in $\mathcal{T}_{prim}$, then do the following. If there is an axiom of the form $(A_1 \sqcap A_2) \sqsubseteq A'$ in $\mathcal{T}_{binc}$, add $\mathbf{G} \cup \{\neg A'\} \setminus \{\neg A_1, \neg A_2\}$ to $\mathcal{T}$. Otherwise, introduce a new internal primitive concept A', add $(\mathbf{G} \cup \{\neg A'\}) \setminus \{\neg A_1, \neg A_2\}$ to $\mathcal{T}$, and add $(A_1 \sqcap A_2) \sqsubseteq A'$ to $\mathcal{T}_{binc}$. Return to Step 1.

   (b) If $\{\neg A_1, \neg A_2\} \subset \mathbf{G}$, and neither A_1 nor A_2 are defined in $\mathcal{T}_{prim}$, then do the following. If there is an axiom of the form $(A_1 \sqcap A_2) \sqsubseteq A'$ in $\mathcal{T}_{binc}$, add $\mathbf{G} \cup \{\neg A'\} \setminus \{\neg A_1, \neg A_2\}$ to $\mathcal{T}$. Otherwise, introduce a new internal primitive concept A', add $(\mathbf{G} \cup \{\neg A'\}) \setminus \{\neg A_1, \neg A_2\}$ to $\mathcal{T}$, and add $(A_1 \sqcap A_2) \sqsubseteq A'$ to $\mathcal{T}_{binc}$. Return to Step 1.

   (c) If $\forall R.C \in \mathbf{G}$, then do the following. Introduce a new internal primitive concept A' and add both $\neg C \sqsubseteq \forall R^{-}.A'$ and $(\mathbf{G} \cup \{\neg A'\}) \setminus \{\forall R.C\}$ to $\mathcal{T}$. Return to Step 1.

   (d) If $\forall R^{-}.C \in \mathbf{G}$, then do the following. Introduce a new internal primitive concept A' and add both $\neg C \sqsubseteq \forall R.A'$ and $(\mathbf{G} \cup \{\neg A'\}) \setminus \{\forall R^{-}.C\}$ to $\mathcal{T}$. Return to Step 1.

4. Try to unfold $\mathbf{G}$. If, for some $A \in \mathbf{G}$ (resp. $\neg A \in \mathbf{G}$), there is an axiom $A \doteq C$ in $\mathcal{T}_{prim}$, then substitute $A \in \mathbf{G}$ (resp. $\neg A \in \mathbf{G}$) with C (resp. $\neg C$), and add $\mathbf{G}$ to $\mathcal{T}$. Return to Step 1.

5. Try to absorb $\mathbf{G}$. If $\neg A \in \mathbf{G}$ and A is not defined in $\mathcal{T}_{prim}$, add $A \sqsubseteq C$ to $\mathcal{T}_{uinc}$ where C is the disjunction of all the concepts in $\mathbf{G} \setminus \{\neg A\}$. Return to Step 1.

6. If none of the above are possible ($\mathbf{G}$ cannot be absorbed), add $\mathbf{G}$ to $\mathcal{T}_g$. Return to Step 1.

Termination of our procedure can be established by a straightforward counting argument involving concept constructors in $\mathcal{T}$. We now prove the correctness of our algorithm using induction. We use the following four lemmas in our induction step. The first two lemmas prove, in combination, that both Step 3(a) and Step 3(b) of our algorithm is correct. The last two lemmas prove Step 3(c) and Step 3(d) correct respectively.

Lemma 4.1. *Let $\mathcal{T}_1, \mathcal{T}_2$, and $\mathcal{T}$ denote TBoxes, $C \in \mathrm{L}$ an arbitrary concept, and A a primitive concept not used in C or $\mathcal{T}$. If $\mathcal{T}_1$ is of the form*

$$\mathcal{T}_1 = \mathcal{T} \cup \{(C_1 \sqcap C_2 \sqcap C_3) \sqsubseteq C_4\},$$

then C is satisfiable with respect to $\mathcal{T}_1$ iff C is satisfiable with respect to

$$\mathcal{T}_2 = \mathcal{T} \cup \{(C_1 \sqcap C_2) \sqsubseteq A, (A \sqcap C_3) \sqsubseteq C_4\}.$$

Proof. First we prove the if direction. Assume C is satisfiable with respect to $\mathcal{T}_1$. For each interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{T}_1$ and $C^{\mathcal{I}} \neq \emptyset$, we extend to an interpretation $\mathcal{I}'$ such that $\mathcal{I}' \models \mathcal{T}_2$ and $C^{\mathcal{I}'} \neq \emptyset$. First, set $\mathcal{I}' = \mathcal{I}$. For each $x \in \Delta^{\mathcal{I}}$ such that $x \in C_1^{\mathcal{I}}$ and $x \in C_2^{\mathcal{I}}$, add x to $A^{\mathcal{I}'}$. Then, $\mathcal{I}' \models \mathcal{T}_2$.

For the only if direction, assume C is satisfiable with respect to $\mathcal{T}_2$. For each interpretation $\mathcal{I} \in \mathrm{Int(L)}$ such that $\mathcal{I} \models \mathcal{T}_2$ and $C^{\mathcal{I}} \neq \emptyset$, we show that $\mathcal{I} \models \mathcal{T}_1$. The proof is by contradiction. Assume $\mathcal{I} \not\models \mathcal{T}_1$. It must be the case that $(C_1 \sqcap C_2 \sqcap C_3) \sqsubseteq C_4 \in \mathcal{T}_1$ does not hold, since the rest of $\mathcal{T}_1$ is a subset of $\mathcal{T}_2$. Therefore, there exists $x \in \Delta^{\mathcal{I}}$ such that $x \in C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}} \cap C_3^{\mathcal{I}}$, and $x \notin C_4^{\mathcal{I}}$. But, in this case either $(C_1 \sqcap C_2) \sqsubseteq A \in \mathcal{T}_2$ or $(A \sqcap C_3) \sqsubseteq C_4 \in \mathcal{T}_2$ must not hold. A contradiction. □

The following lemma proves that instead of introducing a new primitive concept every time we execute Step 3(a) (or Step 3(b)) of our algorithm, we may instead reuse a previously introduced primitive concept. We use **H** to denote an arbitrary axiom.

Lemma 4.2. *Let $\mathcal{T}_1, \mathcal{T}_2$, and $\mathcal{T}$ denote TBoxes, $C \in \mathrm{L}$ an arbitrary concept, A a primitive concept not used in C or $\mathcal{T}$, and A_1, A_2, primitive concepts introduced by Step 3(a) (or Step 3(b)) of our algorithm modified such that a new primitive is always introduced. If $\mathcal{T}_1$ is of the form*

$$\mathcal{T}_1 = \mathcal{T} \cup \{(C_1 \sqcap C_2) \sqsubseteq A_1, (C_1 \sqcap C_2) \sqsubseteq A_2\},$$

then C is satisfiable with respect to $\mathcal{T}_1$ iff C is satisfiable with respect to

$$\mathcal{T}_2 = \{(C_1 \sqcap C_2) \sqsubseteq A\} \cup \{\mathbf{H} \textit{ where } A \textit{ is substituted for } A_1 \textit{ and } A_2 \ \mid \mathbf{H} \in \mathcal{T}\}.$$

Proof. First we prove the if direction. We have two cases.

- Let $\mathcal{I}$ be an interpretation such that $\mathcal{I} \models \mathcal{T}_1$, $C^{\mathcal{I}} \neq \emptyset$, and $A_1^{\mathcal{I}} = A_2^{\mathcal{I}}$. We construct an interpretation $\mathcal{I}'$ from $\mathcal{I}$ such that $\mathcal{I}' \models \mathcal{T}_2$ and $C^{\mathcal{I}'} \neq \emptyset$. First, set $\mathcal{I}' = \mathcal{I}$. Then, set $A^{\mathcal{I}'} = A_1^{\mathcal{I}}$ and remove any references to A_1 and A_2 in $\mathcal{I}'$.

- Let $\mathcal{I}$ be an interpretation such that $\mathcal{I} \models \mathcal{T}_1$, $C^{\mathcal{I}} \neq \emptyset$, and $A_1^{\mathcal{I}} \neq A_2^{\mathcal{I}}$. We construct an interpretation $\mathcal{I}'$ from $\mathcal{I}$ such that $\mathcal{I}' \models \mathcal{T}_1$, $C^{\mathcal{I}'} \neq \emptyset$, and $A_1^{\mathcal{I}} = A_2^{\mathcal{I}}$. For $x \in \Delta^{\mathcal{I}}$ such that $x \in A_1^{\mathcal{I}} \cup A_2^{\mathcal{I}}$ and $x \notin A_1^{\mathcal{I}} \cap A_2^{\mathcal{I}}$, we show that we can remove x from either $A_1^{\mathcal{I}}$ or $A_2^{\mathcal{I}}$ so that $x \notin A_1^{\mathcal{I}} \cup A_2^{\mathcal{I}}$ without causing any axiom in $\mathcal{T}_1$ to fail to hold. Without loss of generality, assume $x \in A_1^{\mathcal{I}}$ and $x \notin A_2^{\mathcal{I}}$. If $x \in C_1$ and $x \in C_2$, then we have a contradiction. Otherwise, we remove x from $A_1^{\mathcal{I}}$. Since either $x \notin C_1^{\mathcal{I}}$ or $x \notin C_2^{\mathcal{I}}$, the axiom $(C_1 \sqcap C_2) \sqsubseteq A_1$ holds. No other axiom in $\mathcal{T}_1$ has A_1 on the right hand side, therefore removing x from $A_1^{\mathcal{I}}$ does not cause any other axiom to fail to hold. Since the above is true for all x such that x is in only one of $A_1^{\mathcal{I}}$ and $A_2^{\mathcal{I}}$, we may remove individuals from $A_1^{\mathcal{I}}$ and $A_2^{\mathcal{I}}$ until $A_1^{\mathcal{I}} = A_2^{\mathcal{I}}$. Then the first case applies.

Now we prove the only if direction. Let $\mathcal{I}$ be an interpretation such that $\mathcal{I} \models \mathcal{T}_2$ and $C^{\mathcal{I}} \neq \emptyset$. We construct an interpretation $\mathcal{I}'$ from $\mathcal{I}$ such that $\mathcal{I}' \models \mathcal{T}_1$ and $C^{\mathcal{I}'} \neq \emptyset$. First set $\mathcal{I}' = \mathcal{I}$. Then, set $A_1^{\mathcal{I}'} = A_2^{\mathcal{I}'} = A^{\mathcal{I}}$. Due to the construction of $\mathcal{T}_2$, $\mathcal{I}' \models \mathcal{T}_1$ and $C^{\mathcal{I}'} \neq \emptyset$. □

Lemma 4.3. *Let $\mathcal{T}_1, \mathcal{T}_2$, and $\mathcal{T}$ denote TBoxes, $C \in$ L an arbitrary concept, A a primitive concept not used in C or $\mathcal{T}$, and R a role. If $\mathcal{T}_1$ is of the form*

$$\mathcal{T}_1 = \mathcal{T} \cup \{\exists R.C_1 \sqsubseteq C_2\},$$

then C is satisfiable with respect to $\mathcal{T}_1$ iff C is satisfiable with respect to TBox

$$\mathcal{T}_2 = \mathcal{T} \cup \{C_1 \sqsubseteq \forall R^-.A, A \sqsubseteq C_2\}.$$

Proof. First we prove the if direction. Assume C is satisfiable with respect to $\mathcal{T}_1$. For an interpretation $\mathcal{I} \in \mathrm{Int}(\mathrm{L})$ such that $\mathcal{I} \models \mathcal{T}_1$ and $C^{\mathcal{I}} \neq \emptyset$, we extend $\mathcal{I}$ to an interpretation $\mathcal{I}'$ such that $\mathcal{I}' \models \mathcal{T}_2$ and $C^{\mathcal{I}'} \neq \emptyset$. First set $\mathcal{I}' = \mathcal{I}$. For each $x \in \Delta^{\mathcal{I}}$ such that $x \in (\exists R.C_1)^{\mathcal{I}} \cap C_2^{\mathcal{I}}$, we add x to $A^{\mathcal{I}'}$. Then, $\mathcal{I}' \models \mathcal{T}_2$ and $C^{\mathcal{I}'} \neq \emptyset$.

Now we prove the only if direction. Assume C is satisfiable with respect to $\mathcal{T}_2$. For each interpretation $\mathcal{I} \in \mathrm{Int}(\mathrm{L})$ such that $\mathcal{I} \models \mathcal{T}_2$ and $C^{\mathcal{I}} \neq \emptyset$, it is also the case that $\mathcal{I} \models \mathcal{T}_1$. The proof is by contradiction. Assume $\mathcal{I} \not\models \mathcal{T}_1$. It must be the case that axiom $\exists R.C_1 \sqsubseteq C_2$ does not hold as all other axioms in $\mathcal{T}_1$ are also in $\mathcal{T}_2$. Then there exists $x \in \Delta^{\mathcal{I}}$ such that $x \in (\exists R.C_1)^{\mathcal{I}}$ and $x \notin C_2^{\mathcal{I}}$. However, this implies that there exists $y \in \Delta^{\mathcal{I}}$ such that $(x, y) \in R^{\mathcal{I}}$ and $y \in C_1$. From axiom $C_1 \sqsubseteq \forall R^-.A$, it must be the case that $x \in A^{\mathcal{I}}$. From axiom $A \sqsubseteq C_2$, it must be the case that $x \in C_2^{\mathcal{I}}$. A contradiction. □

Lemma 4.4. *Let $\mathcal{T}_1, \mathcal{T}_2$, and $\mathcal{T}$ denote TBoxes, $C \in$ L an arbitrary concept, A a primitive concept not used in C or $\mathcal{T}$, and R a role. If $\mathcal{T}_1$ is of the form*

$$\mathcal{T}_1 = \mathcal{T} \cup \{\exists R^-.C_1 \sqsubseteq C_2\},$$

then C is satisfiable with respect to $\mathcal{T}_1$ iff C is satisfiable with respect to TBox

$$\mathcal{T}_2 = \mathcal{T} \cup \{C_1 \sqsubseteq \forall R.A, A \sqsubseteq C_2\}.$$

The proof of this lemma is similar to that of Lemma 4.3.

Theorem 4.1. *For any TBox $\mathcal{T}$, the binary absorption algorithm computes a correct absorption of $\mathcal{T}$.*

Proof. The proof is by induction on the number of iterations of our algorithm. We define an iteration to end when we return to Step 1. We abbreviate the pair

$$(\{A \sqsubseteq C, \neg A \sqsubseteq \neg C \mid A \doteq C \in \mathcal{T}_{prim}\} \cup \mathcal{T}_{uinc} \cup \mathcal{T}_{binc}, \mathcal{T}_g \cup \mathcal{T})$$

as $\mathcal{T}$ and claim that this pair is always a correct binary absorption.

Initially, $\mathcal{T}_{uinc}$, $\mathcal{T}_{binc}$, and $\mathcal{T}_g$ are empty, primitive axioms are in $\mathcal{T}_{prim}$, and the remaining axioms are in $\mathcal{T}$. By Lemma 3.1, Lemma 3.2, Lemma 3.3, and Lemma 3.4, $\mathcal{T}$ is a correct binary absorption at the start of our algorithm.

Assume we just finish iteration i and now perform iteration $i+1$. By our induction hypothesis, $\mathcal{T}$ is a correct binary absorption. We have a number of possible cases.

- If we perform Step 3(a) or Step 3(b) then iteration $i + 1$ is finished. Due to the ordering of Step 3(a) and Step 3(b), newly introduced primitive concepts are absorbed first. Therefore, a newly introduced primitive concept only appears on the right hand side of an axiom once and Lemma 4.1 and Lemma 4.2 apply. We conclude that $\mathcal{T}$ is a correct binary absorption.
- If we perform Step 3(c), then iteration $i + 1$ is finished and by Lemma 4.3, $\mathcal{T}$ is a correct binary absorption.
- If we perform Step 3(d), then iteration $i + 1$ is finished and by Lemma 4.4, $\mathcal{T}$ is a correct binary absorption.
- If we perform any of Steps 1, 2, 5, or 6, then $\mathcal{T}$ is a correct binary absorption at the end of iteration $i + 1$. This is because Steps 1, 2, 5, and 6 use only equivalence preserving operations.

After the final iteration of our algorithm, $\mathcal{T}$ is a correct binary absorption by mathematical induction. $\square$

5 Summary

We have proposed a simple and straightforward generalization of the absorption theory and algorithms pioneered by Horrocks and Tobies [6, 7]. Called *binary absorption*, the basic idea is to allow terminological axioms of the form

$$(A_1 \sqcap A_2) \sqsubseteq C$$

to qualify for lazy unfolding in model building satisfaction procedures for description logics. An important issue for future work is to evaluate the efficacy of our binary absorption algorithm on real-world problems. Our immediate plans in this direction are to measure the reductions in the number of disjunctions in comparison to (basic) absorption for publicly accessible OWL DL ontologies.

References

[1] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications.* Cambridge Univserity Press, 2003.

[2] F. Baader, E. Franconi, B. Hollunder, B. Nebel, and H.-J. Profitlich. An empirical analysis of optimization techniques for terminological representation systems, or: Making KRIS get a move on. *Applied Artificial Intelligence*, 4:109–132, 1994.

[3] E. D. Demaine, A. Lopez-Ortiz, and I. Munro. Adaptive set intersections, unions, and differences. In *SODA '00: Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, pages 743–752, Philadelphia, PA, USA, 2000. Society for Industrial and Applied Mathematics.

[4] A. Deutsch, L.Popa, and V. Tannen. Query reformulation with constraints. *ACM SIGMOD Record*, 35(1):65–73, 2006.

[5] I. Horrocks. Using an Expressive Description Logic: FaCT or Fiction? In A. G. Cohn, L. Schubert, and S. C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference (KR'98)*, pages 636–647. Morgan Kaufmann Publishers, San Francisco, California, 1998.

[6] I. Horrocks and S. Tobies. Optimisation of terminological reasoning. In *Proc. of the 2000 Description Logic Workshop (DL 2000)*, pages 183–192, 2000.

[7] I. Horrocks and S. Tobies. Reasoning with axioms: Theory and practice. In *Proc. of the 7th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2000)*, pages 285–296, 2000.

[8] D. Tsarkov and I. Horrocks. Efficient reasoning with range and domain constraints. In Volker Haarslev and Ralf Möller, editors, *Description Logics*, volume 104 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2004.

[9] J. A. P. Villa. CGU: A common graph utility for DL reasoning and conjunctive query optimization. Master's thesis, David R. Cheriton School of Computer Science, University of Waterloo, 2005.