Decidability of $\mathcal{SHIQ}$ with Complex Role Inclusion Axioms

Ian Horrocks
Department of Computer Science,
University of Manchester, UK
horrocks@cs.man.ac.uk

Ulrike Sattler
Institut für Theoretische Informatik,
TU Dresden, Germany
sattler@tcs.inf.tu-dresden.de

Abstract

Motivated by medical terminology applications, we investigate the decidability of the well known expressive DL, $\mathcal{SHIQ}$, extended with role inclusion axioms (RIAs) of the form $R \circ S \sqsubseteq P$. We show that this extension is undecidable even when RIAs are restricted to the forms $R \circ S \sqsubseteq R$ or $S \circ R \sqsubseteq R$, but that decidability can be regained by further restricting RIAs to be acyclic. We present a tableau algorithm for this DL and report on its implementation, which behaves well in practise and provides important additional functionality in a medical terminology application.

1 Motivation

The description logic (DL) $\mathcal{SHIQ}$ [Horrocks *et al.*, 1999; Horrocks and Sattler, 2002b] is an expressive knowledge representation formalism that extends $\mathcal{ALC}$ [Schmidt-Schauß and Smolka, 1991] (a notational variant of the multi modal logic **K** [Schild, 1991]) with qualifying number restrictions, inverse roles, role inclusion axioms (RIAs) $R \sqsubseteq S$, and transitive roles. The development of $\mathcal{SHIQ}$ was motivated by several applications, one of which was the representation of knowledge about complex physically structured domains found, e.g., in chemical engineering [Sattler, 2000] and medical terminology [Rector and Horrocks, 1997].

Although $\mathcal{SHIQ}$ allows many important properties of such domains to be captured (e.g., transitive and inverse roles), one extremely useful feature that it cannot express is the "propagation" of one property along another property [Padgham and Lambrix, 1994; Rector, 2002; Spackman, 2000]. E.g., it may be useful to express the fact that certain locative properties are transfered across certain partonomic properties so that a trauma or lesion located in a part of a body structure is recognised as being located in the body structure as a whole. This enables highly desirable inferences such as a fracture of the neck of the femur being inferred to be a kind of fracture of the femur, or an ulcer located in the gastric mucosa being inferred to be a kind of stomach ulcer.

The importance of these kinds of inference, particularly in medical terminology applications, is illustrated by the fact that the Grail DL [Rector *et al.*, 1997], which was specifically designed for use with medical terminology, is able to represent these kinds of propagation (although it is quite weak in other respects). Moreover, in another medical terminology application using the comparatively inexpressive DL $\mathcal{ALC}$, a rather complex "work around" is performed in order to represent similar propagations [Schulz and Hahn, 2001].[1] Similar expressiveness was also provided in the CycL language by the `transfersThro` statement [Lenat and Guha, 1989].

It is quite straightforward to extend $\mathcal{SHIQ}$ so that this kind of propagation can be expressed: simply allow for role inclusion axioms of the form $R \circ S \sqsubseteq P$, which then enforces all models $\mathcal{I}$ to interpret the composition of $R^{\mathcal{I}}$ with $S^{\mathcal{I}}$ as a sub-relation of $P^{\mathcal{I}}$. E.g., the above examples translate into

$$\texttt{hasLocation} \circ \texttt{isDivisionOf} \sqsubseteq \texttt{hasLocation},$$

which implies that

$$\texttt{Fracture} \sqcap \exists \texttt{hasLocation}.(\texttt{Neck} \sqcap \exists \texttt{isDivisionOf}.\texttt{Femur})$$

is subsumed by/a specialization of

$$\texttt{Fracture} \sqcap \exists \texttt{hasLocation}.\texttt{Femur}$$

Unfortunately, this extension leads to the undecidability of the interesting inference problems; see [Wessel, 2001] for an undecidability proof and [Baldoni, 1998; Baldoni *et al.*, 1998; Demri, 2001] for the closely related family of *Grammar Logics*. On closer inspection of the problem, we observe that only RIAs of the form $R \circ S \sqsubseteq S$ or $S \circ R \sqsubseteq S$ are required in order to express propagation. Surprisingly, it turns out that $\mathcal{SHIQ}$ extended with this restricted form of RIAs is still undecidable. Decidability can be regained, however, by further restricting the set of RIAs to be *acyclic* (in a non-standard way). This additional restriction does not seem too severe: the above examples are still covered, acyclic sets of RIAs should suffice for many applications, and cycles in RIAs may even be an indicator of modelling flaws [Rector, 2002]. We call this decidable logic $\mathcal{RIQ}$.

Here, we present the above undecidability result and prove the decidability of $\mathcal{SHIQ}$ with acyclic RIAs via a tableau-based decision procedure for concept satisfiability. The algorithm works by transforming concepts of the form $\forall R.C$, where R is a role, into concepts of the form $\forall \mathcal{A}.C$, where $\mathcal{A}$ is a non-deterministic finite automaton (NFA). These automata

[1] In this approach, so-called *SEP-triplets* are used both to compensate for the absence of transitive roles in $\mathcal{ALC}$, and to express the propagation of properties across a distinguished "part-of" role.

are derived from a set of RIAs $\mathcal{R}$ by first *unfolding* $\mathcal{R}$ into a set of implications $\exp(\mathcal{R})$ between regular expressions and roles, and then transforming the regular expressions into automata. The algorithm is of the same complexity as the one for $\mathcal{SHIQ}$—in the size of $\exp(\mathcal{R})$ and the length of the input concept—but, unfortunately, $\exp(\mathcal{R})$ is exponential in $\mathcal{R}$. We present a syntactic restriction that avoids this blow-up; investigating whether this blow-up can be avoided in general will be part of future work. Finally, in order to evaluate the practicability of this algorithm, we have extended the DL system FaCT [Horrocks, 1998] to deal with acyclic RIAs. We discuss how the properties of NFAs are exploited in the implementation, and we present some preliminary results showing that the performance of the extended system is comparable with that of the original, and that it is able to compute inferences of the kind mentioned above w.r.t. the well known Galen medical terminology knowledge base [Rector and Horrocks, 1997; Horrocks, 1998].

For full proofs, the interested reader is referred to [Horrocks and Sattler, 2002a].

2 Preliminaries

In this section, we introduce the DL $\mathcal{SH}^+\mathcal{IQ}$. This includes the definition of syntax, semantics, and inference problems.

Definition 1 *Let* $\mathbf{C}$ *and* $\mathbf{R}$ *be sets of* concept *and* role names. *The set of* roles *is* $\mathbf{R} \cup \{R^- \mid R \in \mathbf{R}\}$. *For roles* R_i *(each of which can be inverse), a* role inclusion axiom (RIA) *is an expression of the form* $R_1 \sqsubseteq R_2$, $R_1 \circ R_2 \sqsubseteq R_1$, *or* $R_1 \circ R_2 \sqsubseteq R_2$. *A* generalised role box (g-RBox) *is a set of RIAs.*

An interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ *associates, with each role name* R, *a binary relation* $R^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. *Inverse roles are interpreted as usual, i.e.,*

$$(R^-)^\mathcal{I} = \{\langle y, x\rangle \mid \langle x, y\rangle \in R^\mathcal{I}\} \quad \textit{for each role } R \in \mathbf{R}.$$

An interpretation $\mathcal{I}$ *is a* model *of a g-RBox* $\mathcal{R}$ *if it satisfies each inclusion assertion in* $\mathcal{R}$, *i.e., if*

$$\begin{array}{rcll} R_1^\mathcal{I} & \subseteq & R_2^\mathcal{I} & \textit{for each } R_1 \sqsubseteq R_2 \in \mathcal{R} \textit{ and} \\ R_1^\mathcal{I} \circ R_2^\mathcal{I} & \subseteq & R_3^\mathcal{I} & \textit{for each } R_1 \circ R_2 \sqsubseteq R_3 \in \mathcal{R}, \end{array}$$

where $\circ$ *stands for the composition of binary relations.*

Transitive role names were not introduced since $R \circ R \sqsubseteq R$ is equivalent to saying that R is a transitive role.

To avoid considering roles such as R^{--}, we define a function Inv on roles as follows: $\mathsf{Inv}(R) = R^-$ if R is a role name, and $\mathsf{Inv}(R) = S$ if $R = S^-$.

Obviously, if $S \circ R \sqsubseteq S \in \mathcal{R}$ ($R \circ S \sqsubseteq S \in \mathcal{R}$ or $R \sqsubseteq S \in \mathcal{R}$), then each model of $\mathcal{R}$ also satisfies $\mathsf{Inv}(R) \circ \mathsf{Inv}(S) \sqsubseteq \mathsf{Inv}(S)$ ($\mathsf{Inv}(S) \circ \mathsf{Inv}(R) \sqsubseteq \mathsf{Inv}(S)$ and $\mathsf{Inv}(R) \sqsubseteq \mathsf{Inv}(S)$). Thus, in the following, we assume that a g-RBox always contains both "directions" of a RIA.

For a g-RBox $\mathcal{R}$, we define the relation $\underline{\circledast}$ to be the transitive-reflexive closure of $\sqsubseteq$ over $\mathcal{R}$.

Definition 2 *A role* S *is* simple *if it does not have implied sub-roles, i.e., if* $S_1 \circ S_2 \sqsubseteq S_3$ *implies* $S_3 \underline{\circledast} S$ *does not hold.*

The set of $\mathcal{SH}^+\mathcal{IQ}$-concepts *is the smallest set such that (i) every concept name is a concept, and, (ii) if* C, D *are concepts,* R *is a role (possibly inverse),* S *is a simple role (possibly inverse) , and* n *is a nonnegative integer, then* $C \sqcap D$, $C \sqcup D$, $\neg C$, $\forall R.C$, $\exists R.C$, $(\geqslant n S.C)$, *and* $(\leqslant n S.C)$ *are also concepts.*

An interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ *consists of a set* $\Delta^\mathcal{I}$, *called the* domain *of* $\mathcal{I}$, *and a* valuation $\cdot^\mathcal{I}$ *which maps every concept to a subset of* $\Delta^\mathcal{I}$ *and every role to a subset of* $\Delta^\mathcal{I} \times \Delta^\mathcal{I}$ *such that, for all concepts* C, D, *roles* R, S, *and non-negative integers* n, *the following equations are satisfied, where* $\sharp M$ *denotes the cardinality of a set* M:

$$\begin{array}{rl} (\neg C)^\mathcal{I} = & \Delta^\mathcal{I} \setminus C^\mathcal{I}, \\ (C \sqcap D)^\mathcal{I} = & C^\mathcal{I} \cap D^\mathcal{I}, \quad (C \sqcup D)^\mathcal{I} = C^\mathcal{I} \cup D^\mathcal{I}, \\ (\exists R.C)^\mathcal{I} = & \{x \mid \exists y.\langle x, y\rangle \in R^\mathcal{I} \textit{ and } y \in C^\mathcal{I}\}, \\ (\forall R.C)^\mathcal{I} = & \{x \mid \forall y.\langle x, y\rangle \in R^\mathcal{I} \textit{ implies } y \in C^\mathcal{I}\}, \\ (\geqslant n R.C)^\mathcal{I} = & \{x \mid \sharp\{y \mid \langle x, y\rangle \in R^\mathcal{I} \textit{ and } y \in C^\mathcal{I}\} \geqslant n\}, \\ (\leqslant n R.C)^\mathcal{I} = & \{x \mid \sharp\{y \mid \langle x, y\rangle \in R^\mathcal{I} \textit{ and } y \in C^\mathcal{I}\} \leqslant n\}. \end{array}$$

A concept C *is called* satisfiable *w.r.t. a g-RBox* $\mathcal{R}$ *iff there is a model* $\mathcal{I}$ *of* $\mathcal{R}$ *with* $C^\mathcal{I} \neq \emptyset$. *A concept* D subsumes *a concept* C *w.r.t.* $\mathcal{R}$ *(written* $C \sqsubseteq_\mathcal{R} D$*) iff* $C^\mathcal{I} \subseteq D^\mathcal{I}$ *holds for each model* $\mathcal{I}$ *of* $\mathcal{R}$. *For an interpretation* $\mathcal{I}$, *an element* $x \in \Delta^\mathcal{I}$ *is called an* instance *of a concept* C *iff* $x \in C^\mathcal{I}$.

Remarks: number restrictions $(\geqslant n R.C)$ and $(\leqslant n R.C)$ are restricted to *simple* roles (intuitively these are (possibly inverse) roles that are not implied by others) since $\mathcal{SHIQ}$ without this restriction is undecidable [Horrocks *et al.*, 1999].

For DLs that are closed under negation, subsumption and (un)satisfiability can be mutually reduced: $C \sqsubseteq D$ iff $C \sqcap \neg D$ is unsatisfiable, and C is unsatisfiable iff $C \sqsubseteq A \sqcap \neg A$ for some concept name A. It is straightforward to extend these reductions to g-RBoxes and TBoxes. In contrast, the reduction of inference problems w.r.t. a *TBox* to pure concept inference problems (possibly w.r.t. a g-RBox), deserves special care: $\mathcal{SHIQ}$ is expressive enough to *internalise* TBoxes, i.e., to reduce reasoning w.r.t. TBoxes to reasoning without TBoxes [Schild, 1991; Horrocks *et al.*, 1999]. Thus, in the following, we restrict our attention to the satisfiability of $\mathcal{SH}^+\mathcal{IQ}$-concepts.

2.1 Relationship with other formalisms

Grammar logics are a class of propositional multi modal logics where the accessibility relations are "axiomatised" through a grammar [Fariñas del Cerro and Penttonen, 1988]. More precisely, for σ_i, τ_j modal parameters, the production rule $\sigma_1 \ldots \sigma_m \rightarrow \tau_1 \ldots \tau_n$ can be viewed as a notational variant for the RIA $\tau_1 \circ \ldots \circ \tau_n \sqsubseteq \sigma_1 \circ \ldots \circ \sigma_m$. Analogously to the DL case, the semantics of a grammar logic takes into account only those frames/relational structures that "satisfy the grammar".

Now grammars are traditionally organised in (refinements of) the Chomsky hierarchy, which induces a hierarchy of grammar logics, e.g., *context free* grammar logics are those propositional multi modal logics where the accessibility relations can be axiomatised through a *context free* grammar. Unsurprisingly, the expressiveness of the grammars influences the expressiveness of the corresponding grammar logics. It was shown that satisfiability of *regular* grammar logics is ExpTime-complete [Demri, 2001], whereas this problem is undecidable for context free grammar logics [Baldoni, 1998; Baldoni *et al.*, 1998]. The latter result is closely related to the undecidability proof in [Wessel, 2001].

Here, we are concerned with (a) multi modal logics that provide for a converse operator on modal parameters and *graded* modalities (to restrict the number of accessible worlds; see, e.g., [Tobies, 2001]) and (b) a certain sub-class of context-free grammars. In our undecidability proof in Section 3, the main difficulty was to develop a grammar that generates the language $\{(ab)^n(cd)^n \mid n \geq 0\}$ using only productions of the form $R \to RS$ or $R \to SR$.[2] We can construct a "similar" grammar G with $L(G) \cap (ab)^*(cd)^* = \{(ab)^n(cd)^n \mid n \geq 0\}$. The production rules of G are

$$\begin{array}{llllll} D & \to & AD, & A & \to & AC, \\ C & \to & BC, & B & \to & BD, \; A \to a, \dots D \to d. \end{array}$$

Role value maps (RVMs) [Brachman and Schmolze, 1985; Schmidt-Schauss, 1989] are closely related to the RIAs investigated here. RVMs are concepts of the form $R_1 \dots R_m \mathrel{\dot\sqsubseteq} S_1 \dots S_n$, for R_i, S_i roles, whose interpretation $(R_1 \dots R_m \mathrel{\dot\sqsubseteq} S_1 \dots S_n)^{\mathcal{I}}$ is defined as follows:

$$\{x \in \Delta^{\mathcal{I}} \mid (R_1 \dots R_m)^{\mathcal{I}}(x) \subseteq (S_1 \dots S_n)^{\mathcal{I}}(x)\},$$

where $(R_1 \dots R_m)^{\mathcal{I}}(x)$ denotes the set of those $y \in \Delta^{\mathcal{I}}$ that are reachable from x via $R_1^{\mathcal{I}} \circ \dots \circ R_m^{\mathcal{I}}$. Thus the RIA $R \circ S \sqsubseteq T$ is equivalent to saying that each individual must be an instance of $RS \mathrel{\dot\sqsubseteq} T$. The undecidability proof of KL-ONE [Schmidt-Schauss, 1989] also involves RVMs $T \mathrel{\dot\sqsubseteq} RS$, and thus cannot be adapted easily to our logic.

3 $\mathcal{SH}^+\mathcal{IQ}$ is undecidable

Due to the syntactic restriction on RIAs, we were not able to adapt the undecidability proof for $\mathcal{ALC}$ with context-free or linear grammars in [Baldoni, 1998; Baldoni *et al.*, 1998; Demri, 2001], the one for $\mathcal{ALC}$ with role boxes [Wessel, 2001], or the one for KL-ONE [Schmidt-Schauss, 1989] to prove undecidability of $\mathcal{SH}^+\mathcal{IQ}$. In the following, we sketch the reduction of the undecidable domino problem [Berger, 1966] to $\mathcal{SH}^+\mathcal{IQ}$ satisfiability.

Definition 3 *A domino system $\mathcal{D} = (D, H, V)$ consists of a non-empty set of domino types $D = \{D_1, \dots, D_n\}$, and of sets of horizontally and vertically matching pairs $H \subseteq D \times D$ and $V \subseteq D \times D$. The problem is to determine if, for a given $\mathcal{D}$, there exists a* tiling *$t : \mathbb{N} \times \mathbb{N} \to D$ such that for all $m, n \in \mathbb{N}$, $\langle t(m,n), t(m+1,n)\rangle \in H$ and $\langle t(m,n), t(m,n+1)\rangle \in V$.*

For a domino system $\mathcal{D}$, we define a $\mathcal{SH}^+\mathcal{IQ}$-concept $C_{\mathcal{D}}$ and a g-RBox $\mathcal{R}_D$ such that $\mathcal{D}$ has a tiling iff $C_{\mathcal{D}}$ is satisfiable w.r.t. $\mathcal{R}_{\mathcal{D}}$. Due to space limitation, we only present $\mathcal{R}_{\mathcal{D}}$:

$$\begin{array}{l} \{v_i \sqsubseteq y_i,\ v_i \sqsubseteq v,\ h_i \sqsubseteq x_i,\ h_i \sqsubseteq h \mid 0 \leq i \leq 3\} \cup \\ \{\ x^-_{i\oplus 1} y_i \ \sqsubseteq\ y_i, \quad x^-_{i\oplus 1} x_i \ \sqsubseteq\ x^-_{i\oplus 1} \\ \quad y^-_{i\oplus 1} x_i \ \sqsubseteq\ x_i, \quad y^-_{i\oplus 1} y_i \ \sqsubseteq\ y^-_{i\oplus 1} \mid 0 \leq i \leq 3\} \end{array}$$

where $\oplus$ and $\ominus$ denotes addition and subtraction modulo four.

Existential and number restrictions on roles h and v (for the horizontal and vertical neighbours) are used to ensure that a point has at most one vertical and at most one horizontal successor, and that these successors satisfy the horizontal and vertical matching conditions induced by H and V; this, as well as ensuring that each point is associated with exactly one domino type, is standard in domino reductions.

[2] Thanks to Christof Löding at RWTH Aachen!

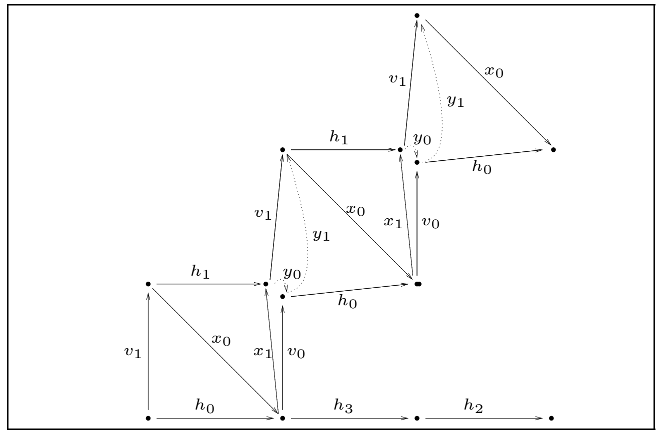

Figure 1: The staircase structure and the effects of $\mathcal{R}_{\mathcal{D}}$.

The next step is rather special: we do not enforce a grid structure, but a structure with "staircases", which is illustrated in Figure 1. To this purpose, we introduce four sub-roles $v_0, \dots, v_3$ of v and four sub-roles $h_0, \dots, h_3$ of h, and ensure that we only have "staircases". An i-staircase is an alternating chain of v_i and h_i edges, without any other v_j- or h_j-successors. At each point on the x-axis, two staircases start that need not meet again, one i-staircase starting with v_i and one $i \ominus 1$-staircase starting with $h_{i\ominus 1}$. A symmetric behaviour is enforced for the nodes on the y-axis.

It only remains to ensure that, if two elements b, b' represent the same point in the grid, then they are associated with the same domino type: b and b' "represent the same point" if there is an n and an instance a on the x-or the y-axis such that both b and b' are reachable by following a staircase starting at a for n steps, i.e., if there is a v_ih_i-path (resp. h_iv_i-path) of length n from a to b, and a $h_{i\ominus 1}v_{i\ominus 1}$-path (resp. $v_{i\oplus 1}h_{i\oplus 1}$-path) of length n from a to b'.

To this purpose, we add super roles x_i of h_i and y_i of v_i (for which we use dashed arrows in Figure 1), and the last group of RIAs in $\mathcal{R}_{\mathcal{D}}$. These role inclusion axioms enforce appropriate, additional role successorships between elements, and we use the additional roles x_i and y_i since we only want to have at most one v_i or h_i-successor. For each 2 staircases starting at the same element on one of the axes, these RIAs ensure that each pair of elements representing the same point is related by y_i. To see this, consider the consequences of the RIAs for elements representing the four points $(1,0), \dots, (2,1)$, and "apply" the RIA $y_1^- x_0 \sqsubseteq x_0$. Next, "apply" $x_1^- x_0 \sqsubseteq x_1^-$, and finally $x_1^- y_0 \sqsubseteq y_0$, which yields the y_0-link between the two elements representing $(2,1)$. Then, starting with $y_1^- y_0 \sqsubseteq y_1$, we can continue with the points $(2,1), \dots, (3,2)$ and work up the role inclusion axioms and up the staircase.

The above observations imply that the concept $C_{\mathcal{D}}$ is satisfiable w.r.t. $\mathcal{T}_{\mathcal{D}}$ and $\mathcal{R}_{\mathcal{D}}$ iff $\mathcal{D}$ has a solution.

Theorem 1 *Satisfiability of $\mathcal{SH}^{+}\mathcal{IQ}$-concepts w.r.t. generalized RBoxes is undecidable.*

4 $\mathcal{RIQ}$ is decidable

In this section, we show that $\mathcal{SHIQ}$ with *acyclic* generalised RBoxes, $\mathcal{RIQ}$, is decidable. We present a tableau-based algorithm that decides satisfiability of $\mathcal{RIQ}$-concepts, and therefore also subsumption in $\mathcal{RIQ}$ and, using internalisation, both inferences w.r.t. TBoxes. The tableau algorithm implemented in the FaCT system [Horrocks, 1998] was extended to the one presented here, and the empirical results are reported in Section 5.

Definition 4 *Let $\mathcal{R}$ be a g-RBox (containing always both directions of a RIA; see above). A role R* directly affects *a role S if $R \neq S$ and either $R \sqsubseteq S \in \mathcal{R}$, $R \circ S \sqsubseteq S \in \mathcal{R}$, or $S \circ R \sqsubseteq S \in \mathcal{R}$. Let* "affects" *be the transitive closure of "directly affects". An* acyclic generalised RBox (a-RBox) *is a g-RBox where "affects" has no cycles. $\mathcal{RIQ}$ is the restriction of $\mathcal{SH}^{+}\mathcal{IQ}$ to a-RBoxes.*

Please note that, in a-RBoxes, we can no longer say that a role R is symmetric using $R \sqsubseteq R^{-}$ and $R^{-} \sqsubseteq R$ since this would yield an "affects" cycle of length 2.

Syntactic transformations Before specifying this algorithm, we transform the RBox to make the presentation of the algorithm easier—basically, we *unfold* the role hierarchy to make all implications explicit.

Firstly, for each (possibly inverse) role R we define two regular expressions as follows:

$$\tau_R \;:=\; \Big(\bigcup_{\substack{S \circ R \sqsubseteq R \in \mathcal{R} \\ S \neq R}} S\Big)^{*} R \Big(\bigcup_{\substack{R \circ T \sqsubseteq R \in \mathcal{R} \\ T \neq R}} T\Big)^{*}$$

$$\rho_R \;:=\; \begin{cases} \tau_R & \text{if } R \circ R \sqsubseteq R \notin \mathcal{R} \\ (\tau_R)^{+} & \text{if } R \circ R \sqsubseteq R \in \mathcal{R} \end{cases}$$

Secondly, we iteratively replace roles in ρ_R with unions of regular expressions of roles, working our way up the affecting relation. We start with roles "almost" minimal w.r.t. the affecting relation, i.e., we start with roles R such that all roles S which affect R are *not* affected. We proceed with roles directly affected by roles that are either already treated or not affected by other roles, and do the following:

$$\rho_R \;:=\; (\rho_R \text{ with } R \text{ replaced with } R \cup \bigcup_{\substack{P \sqsubseteq^{*}_{1} R \\ P \neq R}} \rho_P) \text{ and,}$$

for each $S \neq R$ occurring in ρ_R do

$$\rho_R \;:=\; (\rho_R \text{ with } S \text{ replaced with } \bigcup_{P \sqsubseteq^{*}_{1} S} \rho_P).$$

After this recursion, we define $\exp(\mathcal{R}) := \{\rho_R \sqsubseteq R \mid R \text{ occurs in } \mathcal{R}\}$.

Due to the acyclicity of $\mathcal{R}$, the recursion in this transformation terminates after at most n steps for n the number of role inclusion axioms in $\mathcal{R}$. Please note that, by construction, for each (possibly inverse) role R occurring in $\mathcal{R}$, $\exp(\mathcal{R})$ contains exactly one inclusion $\rho_R \sqsubseteq R$.

For example, for the RIAs $\mathcal{R}$

$$\begin{array}{llllll} R \circ S & \sqsubseteq S, & S \circ W & \sqsubseteq S, & T_1 \circ R_1 & \sqsubseteq R_1, \\ R_2 \circ T_2 & \sqsubseteq R_2, & V \circ T_1 & \sqsubseteq T_1, & R_1 \sqsubseteq R, & R_2 \sqsubseteq R \end{array}$$

the above transformation yield a set $\exp(\mathcal{R})$ containing

$$\begin{array}{rl} (R \cup R_2 T_2^{*} \cup (V^{*}T_1)^{*}R_1)^{*} S W^{*} & \sqsubseteq S, \\ R \cup R_2 T_2^{*} \cup (V^{*}T_1)^{*}R_1 & \sqsubseteq R, \\ R_2 T_2^{*} \sqsubseteq R_2, \quad (V^{*}T_1)^{*}R_1 \sqsubseteq R_1, \quad V^{*}T_1 & \sqsubseteq T_1. \end{array}$$

Unfortunately, the size of $\exp(\mathcal{R})$ can be exponential in the size of $\mathcal{R}$. A further syntactic restriction which avoids this exponential blow-up is described in Section 4.1.

The regular role terms on the left hand side of $\exp(\mathcal{R})$ are read with the standard semantics for regular role expressions, (i.e., using union, composition, and transitive closure of binary relations, see, e.g., [Schild, 1991]). We use $L(\rho)$ to denote the language described by a regular expression ρ.

Lemma 1 *An interpretation $\mathcal{I}$ is a model of an acyclic generalised RBox $\mathcal{R}$ iff $\mathcal{I}$ is a model of $\exp(\mathcal{R})$.*

The Tableau Algorithm tries to construct, for an input $\mathcal{RIQ}$-concept D and an a-RBox $\mathcal{R}$, a tableau (an abstraction of a model) for D w.r.t. $\mathcal{R}$. We can prove that this algorithm constructs a tableau for D and $\mathcal{R}$ iff D is satisfiable w.r.t. $\mathcal{R}$, and thus decides satisfiability of $\mathcal{RIQ}$ concepts w.r.t. an a-RBox. But for the use of NFAs introduced below, this algorithm is quite similar to the one for $\mathcal{SHIQ}$ [Horrocks *et al.*, 1999; Horrocks and Sattler, 2002b].

If R occurs in $\mathcal{R}$, then $\rho_R \sqsubseteq R \in \exp(\mathcal{R})$, and we can build a non-deterministic finite automaton (NFA) $\mathcal{A}^R$ with $L(\mathcal{A}^R) = L(\rho_R)$. Due to the use of non-deterministic automata, $\mathcal{A}^R$ can be of size linear in $|\rho_R|$. Otherwise, $\mathcal{A}^R$ is a (two-state) automaton with $L(\mathcal{A}^R) = \{R\}$.

For $\mathcal{A}$ an NFA and q a state in $\mathcal{A}$, $\mathcal{A}_q$ denotes the NFA obtained from $\mathcal{A}$ by making q the (only) initial state of $\mathcal{A}$, and we use $q \rightarrow_S q' \in \mathcal{A}$ to denote that $\mathcal{A}$ has a transition labelled with S from q to q'.

As usual, each concept can be easily transformed into an equivalent one in negation normal form (NNF, i.e., negation occurs in front of concept names only), and we use $\dot\neg C$ for the NNF of a concept C. For a concept C, $\mathsf{clos}(C)$ is the smallest set that contains C and that is closed under sub-concepts and $\dot\neg$. Then $\mathsf{fclos}(C, \mathcal{R})$ is the superset of $\mathsf{clos}(C, \mathcal{R})$ that contains $\forall \mathcal{A}^S_q.D$ for each S occurring in $\mathcal{R}$ or C with q a state in $\mathcal{A}^S$ and $\forall S.D \in \mathsf{clos}(C)$.

A *completion tree* $\mathbf{T}$ for a $\mathcal{RIQ}$ concept D and an a-RBox $\mathcal{R}$ is a tree where each node x is labelled with a set $\mathcal{L}(x) \subseteq \mathsf{fclos}(D, \mathcal{R})$ and each edge $\langle x, y\rangle$ from a node x to its successor y is labelled with a non-empty set $\mathcal{L}(\langle x, y\rangle) \subseteq \mathcal{R}$ of (possibly inverse) roles occurring in D and $\mathcal{R}$. Finally, completion trees come with an explicit inequality relation $\dot{\neq}$ on nodes which is implicitly assumed to be symmetric.

If $R \in \mathcal{L}(\langle x, y\rangle)$ for a node x and its successor y and $R \sqsubseteq^{*} S$, then y is called an S-*successor* of x and x is called an $\mathsf{Inv}(S)$-*predecessor* of y. If y is an S-successor or an $\mathsf{Inv}(S)$-predecessor of x, then y is called an S-neighbour of x. Finally, *ancestor* is the transitive closure of *predecessor*.

For a role S, a concept C and a node x in $\mathbf{T}$ we define $S^{\mathbf{T}}(x, C) := \{y \mid y \text{ is an } S\text{-neighbour of } x \text{ and } C \in \mathcal{L}(y)\}$.

A node is *blocked* iff it is either directly or indirectly blocked. A node x is *directly blocked* iff none of its ancestors are blocked, and it has ancestors x', y and y' such that (1) y is not the root node; (2) x is a successor of x' and y is a successor of y'; *and* (3) $\mathcal{L}(x) = \mathcal{L}(y)$, $\mathcal{L}(x') = \mathcal{L}(y')$, $\mathcal{L}(\langle x', x\rangle) = \mathcal{L}(\langle y', y\rangle)$. A node y is *indirectly blocked* if one of its ancestors is blocked.

$\mathsf{R}\sqcap$: if $C_1 \sqcap C_2 \in \mathcal{L}(x)$, x is not indirectly blocked,
and $\{C_1, C_2\} \not\subseteq \mathcal{L}(x)$
then $\mathcal{L}(x) := \mathcal{L}(x) \cup \{C_1, C_2\}$

$\mathsf{R}\sqcup$: if $C_1 \sqcup C_2 \in \mathcal{L}(x)$, x is not indirectly blocked,
and $\{C_1, C_2\} \cap \mathcal{L}(x) = \emptyset$
then $\mathcal{L}(x) := \mathcal{L}(x) \cup \{E\}$ for some $E \in \{C_1, C_2\}$

$\mathsf{R}\exists$: if $\exists S.C \in \mathcal{L}(x)$, x is not blocked, and
x has no S-neighbour y with $C \in \mathcal{L}(y)$
then create a new node y with
$\mathcal{L}(\langle x, y\rangle) := \{S\}$ and $\mathcal{L}(y) := \{C\}$

$\mathsf{R}\forall_1$: if $\forall S.C \in \mathcal{L}(x)$, x is not indirectly blocked, and
$\forall \mathcal{A}^S.C \notin \mathcal{L}(x)$
then $\mathcal{L}(x) := \mathcal{L}(x) \cup \{\forall \mathcal{A}^S.C\}$

$\mathsf{R}\forall_2$: if $\forall \mathcal{A}_p.C \in \mathcal{L}(x)$, x is not indirectly blocked,
$p \rightarrow_S q$ in $\mathcal{A}_p$, and y is an S-neighbour of x with
$\forall \mathcal{A}_q.C \notin \mathcal{L}(y)$
then $\mathcal{L}(y) := \mathcal{L}(y) \cup \{\forall \mathcal{A}_q.C\}$

$\mathsf{R}\forall_3$: if $\forall \mathcal{A}.C \in \mathcal{L}(x)$, x is not indirectly blocked,
$\varepsilon \in L(\mathcal{A})$, and $C \notin \mathcal{L}(x)$
then $\mathcal{L}(x) := \mathcal{L}(x) \cup \{C\}$

$\mathsf{R?}$: if $(\leqslant nS.C) \in \mathcal{L}(x)$, x is not indir. blocked, and y
is an S-neighbour of x with $\{C, \dot{\neg}C\} \cap \mathcal{L}(y) = \emptyset$
then $\mathcal{L}(y) := \mathcal{L}(y) \cup \{E\}$ for some $E \in \{C, \dot{\neg}C\}$

$\mathsf{R}\geqslant$: if $(\geqslant nS.C) \in \mathcal{L}(x)$, x is not blocked, and there are
no $y_1,..., y_n \in S^{\mathbf{T}}(x, C)$ with $y_i \neq y_j$ for each $i \neq j$
then create n new nodes y_i with $\mathcal{L}(\langle x, y_i\rangle) = \{S\}$,
$\mathcal{L}(y_i) = \{C\}$, and $y_i \neq y_j$ for $1 \leq i < j \leq n$.

$\mathsf{R}\leqslant$: if $(\leqslant nS.C) \in \mathcal{L}(x)$, x is not indirectly blocked,
$\# S^{\mathbf{T}}(x, C) > n$, there are $y, z \in S^{\mathbf{T}}(x, C)$ with
not $y \neq z$ and y is not an ancestor of z,
then $\mathcal{L}(z) := \mathcal{L}(z) \cup \mathcal{L}(y)$ and if z is an ancestor of x
then $\mathcal{L}(\langle z, x\rangle) := \mathcal{L}(\langle z, x\rangle) \cup \mathsf{Inv}(\mathcal{L}(\langle x, y\rangle))$
else $\mathcal{L}(\langle x, z\rangle) := \mathcal{L}(\langle x, z\rangle) \cup \mathcal{L}(\langle x, y\rangle)$
and remove y and the sub-tree below y

Figure 2: The Expansion Rules for $\mathcal{RIQ}$.

For a node x, $\mathcal{L}(x)$ is said to contain a *clash* if, for some concept name A, $\{A, \neg A\} \subseteq \mathcal{L}(x)$, or if there is some concept $(\leqslant nS.C) \in \mathcal{L}(x)$ and x has $n+1$ S-neighbours $y_0, \ldots, y_n$ with $C \in \mathcal{L}(y_i)$ and $y_i \neq y_j$ for all $0 \leq i < j \leq n$. A completion tree is *clash-free* if none of its nodes contains a clash, and it is *complete* if no rule from Figure 2 can be applied to it.

For a $\mathcal{RIQ}$-concept D, the algorithm starts with the completion tree consisting of a single root node x with $\mathcal{L}(x) = \{D\}$ and $\neq$ empty. It applies the expansion rules in Figure 2, stopping when a clash occurs, and answers "D is satisfiable w.r.t. $\mathcal{R}$" iff the completion rules can be applied in such a way that they yield a complete and clash-free completion tree, and "D is unsatisfiable w.r.t. $\mathcal{R}$" otherwise.

Most of the rules have been used before for fragments of $\mathcal{RIQ}$—only the three $\forall_i$-rules are new: they are elegant generalisations of standard rules for value restrictions taking into account automata.

As usual, we can prove termination, soundness, and completeness of the tableau algorithm to show that it indeed decides satisfiability of $\mathcal{RIQ}$-concepts w.r.t. a-RBoxes.

Theorem 2 *The tableau algorithm decides satisfiability and subsumption of $\mathcal{RIQ}$-concepts w.r.t. a-RBoxes and TBoxes.*

4.1 Avoiding the blow-up

So far, the satisfiability algorithm presented here involves an exponential blow-up compared to similar algorithms that are implemented in state-of-the-art description logic reasoners [Horrocks, 1998; Haarslev and Möller, 2001]: the closure $\mathsf{fclos}(D, \mathcal{R})$ is exponential in $\mathcal{R}$ since we have "unfolded" the a-RBox $\mathcal{R}$ into the possibly exponentially large $\exp(\mathcal{R})$. While investigating whether and how this exponential blow-up can be avoided, we observe that a further restriction of the syntax of a-RBoxes avoids this blow-up:

An a-RBox $\mathcal{R}$ is called *simple* if, whenever $R_1 \circ S \sqsubseteq S$ and $S \circ R_2 \sqsubseteq S$ are in $\mathcal{R}$, then R_1 and R_2 do not have a common subrole R' that occurs on the right hand side of an axiom $R' \circ S' \sqsubseteq R'$ or $S' \circ R' \sqsubseteq R'$.

For a *simple* a-RBox $\mathcal{R}$, $\exp(\mathcal{R})$ is only polynomial in the size of $\mathcal{R}$ since each term used in the substitution step of the construction of $\exp(\mathcal{R})$ from $\mathcal{R}$ is at most used once in each other axiom.

Thus, for simple role hierarchies, the tableau algorithm presented here is of the same worst case complexity as for $\mathcal{SHIQ}$, namely 2NExpTime. A detailed investigation of the exact complexity will be part of future work.

5 Empirical Evaluation

In order to evaluate the practicability of the above algorithm, we have extended the DL system FaCT [Horrocks, 1998] to deal with $\mathcal{RIQ}$, and we have carried out a preliminary empirical evaluation.

¿From a practical point of view, one potential problem with the $\mathcal{RIQ}$ algorithm is that the number of different automata, and hence the number of different $\forall \mathcal{A}.C$ concepts, could be very large. Moreover, many of these automata could be equivalent (i.e., accept the same languages). This could adversely effect blocking, and thus lead to a serious degradation in performance [Horrocks and Sattler, 2002b].

The FaCT implementation addresses these possible problems by transforming all of the initial NFAs into minimal deterministic finite automata (DFAs) using the AT&T FSM Library™ [Mohri *et al.*, 1998]. One DFA is constructed for each role, the states in each automaton are uniquely numbered, and the implementation uses concepts of the form $\forall \mathcal{A}.C$, where $\mathcal{A}$ is the number of a state in one of the automata. Because the automata are deterministic, for each concept of the form $\forall \mathcal{A}.C$ in the label of a node x, the $\mathsf{R}\forall_2$-rule can add at most one concept to the label of a given neighbouring node y per role in the label of the edge $\langle x, y\rangle$. Moreover, because the automata are minimal, if $\forall \mathcal{A}.C$ leads to the presence of $\forall \mathcal{A}'.C$ in some successor node as a result of repeated applications of the $\mathsf{R}\forall_2$-rule, then $\forall \mathcal{A}.C$ is equivalent to $\forall \mathcal{A}'.C$ iff $\mathcal{A} = \mathcal{A}'$. As $\mathcal{A}$ and $\mathcal{A}'$ are numbers, such comparisons are very easy, and minimisation of automata avoids unnecessary blocking delays.

The implementation is still at the "beta" stage, but it has been possible to carry out some preliminary tests using the well known Galen medical terminology KB [Rector and Horrocks, 1997; Horrocks, 1998]. This KB contains 2,740 named concepts and 413 roles, 26 of which are transitive. The roles are arranged in a relatively complex hierarchy with a maximum depth of 10. Classifying this KB using FaCT's $\mathcal{SHIQ}$ reasoner takes 116s on an 800 MHz Pentium III equipped Linux PC. Classifying the same KB using the new $\mathcal{RIQ}$ reasoner took a total of 275s, but this includes 135s to compute the minimal DFAs for the role box (it should be noted that this is an unusually large and complex role box, and that computing the DFAs is a preprocessing step that will not need to be repeated when the remainder of the KB is extended, modified, or queried). This result is encouraging as it shows that, in the case of the Galen KB at least, using automata in $\forall\mathcal{A}.C$ concepts does not, in itself, lead to an unacceptable degradation in performance.

The KB was then extended with several RIAs that express the propagation of location across various partonomic roles. These included

$$\texttt{hasLocation} \circ \texttt{isSolidDivisionOf} \sqsubseteq \texttt{hasLocation},$$
$$\texttt{hasLocation} \circ \texttt{isLayerOf} \sqsubseteq \texttt{hasLocation}.$$

Classifying the extended KB took 280s, an increase of only 2% (3.5% if we exclude the DFA computation time). Subsumption queries w.r.t. this KB revealed that, e.g.,

$\texttt{Fracture} \sqcap \exists\texttt{hasLocation}.\texttt{NeckOfFemur}$
was implicitly a kind of
$\texttt{Fracture} \sqcap \exists\texttt{hasLocation}.\texttt{Femur}$

(NeckOfFemur is a solid division of Femur), and

$\texttt{Ulcer} \sqcap \exists\texttt{hasLocation}.\texttt{GastricMucosa}$
was implicitly a kind of
$\texttt{Ulcer} \sqcap \exists\texttt{hasLocation}.\texttt{Stomach}$

(GastricMucosa is a layer of Stomach). None of these subsumption relationships held w.r.t. the original KB. The times taken to compute these relationships w.r.t. the classified KB could not be measured accurately as they were of the same order as a system clock tick (10ms).

References

[Baldoni *et al.*, 1998] M. Baldoni, L. Giordano, and A. Martelli. A tableau calculus for multimodal logics and some (un)decidability results. In *Proc. of TABLEAUX-98*, vol. 1397 of *LNAI*. Springer-Verlag, 1998.

[Baldoni, 1998] M. Baldoni. *Normal Multimodal Logics: Automatic Deduction and Logic Programming Extension*. PhD thesis, Dip. di Informatica, Univ. degli Studi di Torino, Italy, 1998.

[Berger, 1966] R. Berger. The undecidability of the dominoe problem. *Mem. Amer. Math. Soc.*, 66, 1966.

[Brachman and Schmolze, 1985] R. J. Brachman and J. Schmolze. An overview of the KL-ONE knowledge representation system. *Cognitive Science*, 9(2):171–216, 1985.

[Demri, 2001] S. Demri. The complexity of regularity in grammar logics and related modal logics. *J. of Logic and Computation*, 11(6), 2001.

[Fariñas del Cerro and Penttonen, 1988] L. Fariñas del Cerro and M. Penttonen. Grammar logics. *Logique et Analyse*, 121-122:123–134, 1988.

[Haarslev and Möller, 2001] V. Haarslev and R. Möller. RACER system description. In *IJCAR-01*, vol. 2083 of *LNAI*. Springer-Verlag, 2001.

[Horrocks and Sattler, 2002a] I. Horrocks and U. Sattler. Decidability of $\mathcal{SHIQ}$ with complex role inclusion axioms. LTCS-Report 02-06, Dresden University of Technology, Germany, 2002, available at http://lat.inf.tu-dresden.de/research/reports.html.

[Horrocks and Sattler, 2002b] I. Horrocks and U. Sattler. Optimised reasoning for shiq. In *Proc. of ECAI 2002*, IOS Press, 2002.

[Horrocks *et al.*, 1999] I. Horrocks, U. Sattler, and S. Tobies. Practical reasoning for expressive description logics. In *Proc. of LPAR'99*, vol. 1705 in LNAI. Springer-Verlag, 1999.

[Horrocks, 1998] I. Horrocks. Using an Expressive Description Logic: FaCT or Fiction? In *Proc. of KR-98*, Morgan Kaufmann, 1998.

[Lenat and Guha, 1989] D. B. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems*. Addison Wesley, 1989.

[Mohri *et al.*, 1998] M. Mohri, F. C. N. Pereira, and M. Riley. *A Rational Design for a Weighted Finite-State Transducer Library*. vol. 1436 in LNCS. Springer-Verlag, 1998.

[Padgham and Lambrix, 1994] L. Padgham and P. Lambrix. A framework for part-of hierarchies in terminological logics. In *Proc. of KR-94*, Morgan Kaufmann, 1994.

[Rector and Horrocks, 1997] A. Rector and I. Horrocks. Experience building a large, re-usable medical ontology using a description logic with transitivity and concept inclusions. In *Proc. of the WS on Ontological Engineering, AAAI Spring Symposium (AAAI'97)*. AAAI Press, 1997.

[Rector *et al.*, 1997] A. Rector, S. Bechhofer, C. A. Goble, I. Horrocks, W. A. Nowlan, and W. D. Solomon. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*, 9:139–171, 1997.

[Rector, 2002] A. Rector. Analysis of propagation along transitive roles: Formalisation of the galen experience with medical ontologies. In *Proc. of DL 2002*, CEUR-WS, 2002.

[Sattler, 2000] U. Sattler. Description logics for the representation of aggregated objects. In *Proc. of ECAI 2000*. IOS Press, 2000.

[Schild, 1991] K. Schild. A correspondence theory for terminological logics: Preliminary report. In *Proc. of IJCAI-91*, 1991.

[Schmidt-Schauß and Smolka, 1991] M. Schmidt-Schauß and G. Smolka. Attributive concept descriptions with complements. *Artificial Intelligence*, 48(1):1–26, 1991.

[Schmidt-Schauss, 1989] M. Schmidt-Schauss. Subsumption in KL-ONE is undecidable. In *Proc. of KR-89*, Morgan Kaufmann, 1989.

[Schulz and Hahn, 2001] S. Schulz and U. Hahn. Parts, locations, and holes - formal reasoning about anatomical structures. In *Proc. of AIME 2001*, vol. 2101 of *LNAI*. Springer-Verlag, 2001.

[Spackman, 2000] K. Spackman. Managing clinical terminology hierarchies using algorithmic calculation of subsumption: Experience with SNOMED-RT. *J. of the Amer. Med. Informatics Ass.*, 2000. Fall Symposium Special Issue.

[Tobies, 2001] S. Tobies. PSPACE reasoning for graded modal logics. *J. of Logic and Computation*, 11(1):85–106, 2001.

[Wessel, 2001] M. Wessel. Obstacles on the way to qualitative spatial reasoning with description logics: Some undecidability results. In *Proc. of DL 2001*, CEUR-WS, 2001.