An Extension of Complex Role Inclusion Axioms in the Description Logic $\mathcal{SROIQ}$

Yevgeny Kazakov

Oxford University Computing Laboratory

Abstract. We propose an extension of the syntactic restriction for complex role inclusion axioms in the description logic $\mathcal{SROIQ}$. Like the original restriction in $\mathcal{SROIQ}$, our restrictions can be checked in polynomial time and they guarantee regularity for the sets of role chains implying roles, and thereby decidability for the main reasoning problems. But unlike the original restrictions, our syntactic restrictions can represent any regular compositional properties on roles. In particular, many practically relevant complex role inclusion axioms, such as those describing various parthood relations, can be expressed in our extension, but could not be expressed in the original $\mathcal{SROIQ}$.

1 Introduction

The description logic (DL) $\mathcal{SROIQ}$ [11] provides a logical foundation for the new version of the web ontology language OWL 2.[1] In comparison to the DL $\mathcal{SHOIN}$ which underpins the first version of OWL,[2] $\mathcal{SROIQ}$ provides several new constructors for classes and axioms. One of the new powerful features of $\mathcal{SROIQ}$ are so-called complex role inclusion axioms (RIAs) which allow for expressing implications between role chains and roles: $R_1 \cdots R_n \sqsubseteq R$. It is well-known that unrestricted usage of such axioms can easily lead to undecidability for modal and description logics [6, 8, 9, 12], with a notable exception of the DL $\mathcal{EL}^{++}$ [2]. Therefore, certain syntactic restrictions are imposed on RIAs in $\mathcal{SROIQ}$ to regain decidability. Specifically, the restrictions ensure that RIAs $R_1 \cdots R_n \sqsubseteq R$ when viewed as production rules for context-free grammars $R \rightarrow R_1 \dots R_n$ induce a regular language. In fact, the reasoning procedures for $\mathcal{SROIQ}$ [11, 13] do not use the syntactic restrictions directly, but take as an input the resulting non-deterministic finite automata for RIAs. This means that it is possible to use exactly the same procedure for any set of RIAs for which the corresponding regular automata can be constructed.

Unfortunately, the syntactic restrictions on RIAs in $\mathcal{SROIQ}$ are not satisfied in all cases when the language induced by the RIAs is regular. In this paper we analyze several common use cases of RIAs which correspond to regular languages but cannot be expressed within $\mathcal{SROIQ}$. To extend the expressive power of RIAs, we introduce a notion of stratified set of RIAs and demonstrate that it can be used to express the required axioms. Our restrictions have several nice properties, which could allow for their seamless integration into future revisions of OWL:

[1] http://www.w3.org/TR/owl2-overview/

[2] http://www.w3.org/TR/owl-ref/

Table 1. The syntax and semantics of $\mathcal{SROIQ}$

Name	Syntax	Semantics
Concepts		
atomic concept	A	$A^{\mathcal{I}}$ (given)
nominal	$\{a\}$	$\{a^{\mathcal{I}}\}$
top concept	$\top$	$\Delta^{\mathcal{I}}$
negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists R.C$	$\{x \mid R^{\mathcal{I}}(x, C^{\mathcal{I}}) \neq \emptyset\}$
min cardinality	$\geqslant nS.C$	$\{x \mid \|S^{\mathcal{I}}(x, C^{\mathcal{I}})\| \geq n\}$
exists self	$\exists S.\mathsf{Self}$	$\{x \mid \langle x, x\rangle \in S^{\mathcal{I}}\}$
Axioms		
complex role inclusion	$\boldsymbol{\rho \sqsubseteq R}$	$\boldsymbol{\rho^{\mathcal{I}} \subseteq R^{\mathcal{I}}}$
disjoint roles	$\mathsf{Disj}(S_1, S_2)$	$S_1^{\mathcal{I}} \cap S_2^{\mathcal{I}} = \emptyset$
concept inclusion	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
concept assertion	$C(a)$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$
role assertion	$R(a, b)$	$\langle a, b\rangle \in R^{\mathcal{I}}$

1. Our restrictions are *conservative* over the current restrictions in $\mathcal{SROIQ}$. That is, every set of RIAs that satisfies the current restriction in $\mathcal{SROIQ}$ will automatically satisfy our restrictions.
2. Our restrictions are *tractable*, that is, they can be verified in polynomial time in the size of the input set of RIAs.
3. Our restrictions are *constructive*, which means that there is a procedure that builds the corresponding regular automaton for every set of RIAs that satisfies our restrictions.
4. Finally, unlike the original restrictions in $\mathcal{SROIQ}$, our restrictions are *complete* w.r.t. regular compositional properties. This means that any regular compositional properties on roles can be expressed using a stratified set of RIAs.

2 Preliminaries

In this section we introduce syntax and semantics of the DL $\mathcal{SROIQ}$ [11]. A $\mathcal{SROIQ}$ *vocabulary* consists of countably infinite sets N_C of *atomic concepts*, N_R of *atomic roles*, and N_I of *individuals*. A $\mathcal{SROIQ}$ *role* is either $r \in \mathsf{N}_R$, an *inverse role* r^- with $r \in \mathsf{N}_R$, or the *universal role* U. A *role chain* is a sequence of roles $\rho = R_1 \cdots R_n$, $n \geq 0$, where $R_i \neq U$, $1 \leq i \leq n$; in this case we denote by $\|\rho\| := n$ the *size* of ρ; when $n = 0$, ρ is called the *empty role chain* and is denoted by ϵ. With $\rho_1\rho_2$ we denote the *concatenation* of role chains ρ_1 and ρ_2, and with ρR ($R\rho$) we denote the role chain obtained by appending (prepending) R to ρ. We denote by $\mathsf{Inv}(R)$ the *inverse of a role* R defined by $\mathsf{Inv}(R) := r^-$ when $R = r$, $\mathsf{Inv}(R) := r$ when $R = r^-$, and $\mathsf{Inv}(R) := U$ when $R = U$. The *inverse of a role chain* $\rho = R_1 \cdots R_n$ is a role chain $\mathsf{Inv}(\rho) := \mathsf{Inv}(R_n) \cdots \mathsf{Inv}(R_1)$.

The syntax and semantics of $\mathcal{SROIQ}$ is summarized in Table 1. The set of $\mathcal{SROIQ}$

concepts is recursively defined using the constructors in the upper part of the table, where $A \in \mathsf{N}_C$, C, D are concepts, R, S roles, a an individual, and n a positive integer.

A *regular order on roles* is an irreflexive transitive binary relation $\prec$ on roles such that $R_1 \prec R_2$ iff $\mathsf{Inv}(R_1) \prec R_2$. A (complex) *role inclusion axiom* (RIA) $R_1 \cdots R_n \sqsubseteq R$ is said to be *$\prec$-regular*, if either: (i) $n = 2$ and $R_1 = R_2 = R$, or (ii) $n = 1$ and $R_1 = \mathsf{Inv}(R)$, or (iii) $R_i \prec R$ for $1 \leq i \leq n$, or (iv) $R_1 = R$ and $R_i \prec R$ for $1 < i \leq n$, or (v) $R_n = R$ and $R_i \prec R$ for $1 \leq i < n$. A set $\mathcal{R}$ of RIAs is *$\prec$-regular* if every RIA in $\mathcal{R}$ is $\prec$-regular.

A $\mathcal{SROIQ}$ *ontology* is a set $\mathcal{O}$ of axioms listed in the lower part of Table 1, where ρ is a role chain, $R_{(i)}$ and $S_{(i)}$ are roles, C, D concepts, and a, b individuals, such that the set of all RIAs in $\mathcal{O}$ is $\prec$-regular for some regular order $\prec$ on roles.

For a RIA $\alpha = (\rho \sqsubseteq R)$ and role chains ρ' and ρ'', we write $\rho' \sqsubseteq_\alpha \rho''$ if $\rho' = \rho'_1 \rho \rho'_2$ and $\rho'' = \rho'_1 R \rho'_2$ for some ρ'_1 and ρ'_2. To indicate a position where α was used, we also write $\rho' \sqsubseteq_{\alpha,k} \rho''$ where $k = \|\rho'_1\|$. For a set of RIAs $\mathcal{R}$, we write $\rho' \sqsubseteq_\mathcal{R} \rho''$ ($\rho' \sqsubseteq_{\mathcal{R},k} \rho''$) if $\rho' \sqsubseteq_\alpha \rho''$ ($\rho' \sqsubseteq_{\alpha,k} \rho''$) for some $\alpha \in \mathcal{R}$. We denote by $\sqsubseteq^*_\mathcal{R}$ ($\sqsubseteq^*_{\mathcal{R},k}$) the reflexive transitive closure of $\sqsubseteq_\mathcal{R}$ ($\sqsubseteq_{\mathcal{R},k}$). The sequence $\rho_0 \sqsubseteq_{\alpha_1} \rho_1 \cdots \sqsubseteq_{\alpha_n} \rho_n$ ($\rho_0 \sqsubseteq_{\alpha_1,k_1} \rho_1 \cdots \sqsubseteq_{\alpha_n,k_n} \rho_n$), $n \geq 0$, $\alpha_i \in \mathcal{R}$ $(1 \leq i \leq n)$ is called a *proof for* $\rho_0 \sqsubseteq \rho_n$ *in* $\mathcal{R}$. In this case we also say that $\rho_0 \sqsubseteq \rho_n$ *is provable in* $\mathcal{R}$.

We denote by $\bar{\mathcal{R}}$ the extension of $\mathcal{R}$ with *inverses* $\mathsf{Inv}(\rho) \sqsubseteq \mathsf{Inv}(R)$ of RIAs $\rho \sqsubseteq R \in \mathcal{R}$. Let $\mathcal{O}$ be a $\mathcal{SROIQ}$ ontology and $\mathcal{R}$ the set of RIAs in $\mathcal{O}$. A role S is *simple* if $\rho \sqsubseteq^*_{\bar{\mathcal{R}}} S$ implies $\|\rho\| \leq 1$. It is required that all roles $S_{(i)}$ in Table 1 are simple w.r.t. $\mathcal{R}$. Other constructors of $\mathcal{SROIQ}$ [11] can be expressed using those in Table 1. The *bottom concept* $\bot$ stands for $\neg\top$, *disjunction* $C \sqcup D$ for $\neg(\neg C \sqcap \neg D)$, *universal restriction* $\forall R.C$ for $\neg(\exists R.\neg C)$, *max cardinality* $\leqslant n S.C$ for $\neg(\geqslant (n+1) S.C)$, *role transitivity* $\mathsf{Tra}(S)$ for $S \cdot S \sqsubseteq S$, *role reflexivity* $\mathsf{Ref}(R)$ for $\epsilon \sqsubseteq R$, *role symmetry* $\mathsf{Sym}(R)$ for $\mathsf{Inv}(R) \sqsubseteq R$, *role irreflexivity* $\mathsf{Irr}(S)$ for $\exists S.\mathsf{Self} \sqsubseteq \bot$, *role asymmetry* $\mathsf{Asy}(S)$ for $\mathsf{Disj}(S, \mathsf{Inv}(S))$, *concept equivalence* $C \equiv D$ for $C \sqsubseteq D$ and $D \sqsubseteq C$, and *negative role assertion* $\neg S(a, b)$ for $S_S(a, b)$ and $\mathsf{Disj}(S, S_S)$, where S_S is a fresh (simple) role for S. Of all constructors and axioms, only RIAs are of a primary focus in this paper.

The semantics of $\mathcal{SROIQ}$ is defined using interpretations. An *interpretation* is a pair $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ where $\Delta^\mathcal{I}$ is a non-empty set called *the domain of the interpretation* and $\cdot^\mathcal{I}$ is the *interpretation function*, which assigns to every $A \in \mathsf{N}_C$ a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, to every $r \in \mathsf{N}_R$ a relation $r^\mathcal{I} \in \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, and to every $a \in \mathsf{N}_I$ an element $a^\mathcal{I} \in \Delta^\mathcal{I}$. $\mathcal{I}$ is extended to roles by $U^\mathcal{I} := \Delta^\mathcal{I} \times \Delta^\mathcal{I}$ and $(r^-)^\mathcal{I} := \{\langle x, y\rangle \mid \langle y, x\rangle \in r^\mathcal{I}\}$, and to role chains by $(R_1 \cdots R_n)^\mathcal{I} := R_1^\mathcal{I} \circ \cdots \circ R_n^\mathcal{I}$ where $\circ$ is the composition of binary relations. The empty role chain ϵ is interpreted by $\epsilon^\mathcal{I} := \{\langle x, x\rangle \mid x \in \Delta^\mathcal{I}\}$.

The interpretation of concepts is defined according to the right column of the upper part of Table 1, where $\delta(x, V)$ for $\delta \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, $V \subseteq \Delta^\mathcal{I}$, and $x \in \Delta^\mathcal{I}$ denotes the set $\{y \mid \langle x, y\rangle \in \delta \wedge y \in V\}$, and $\|V\|$ denotes the cardinality of a set $V \subseteq \Delta^\mathcal{I}$. An interpretation $\mathcal{I}$ *satisfies an axiom* α (written $\mathcal{I} \models \alpha$) if the respective condition to the right of the axiom in Table 1 holds; $\mathcal{I}$ is a *model of an ontology* $\mathcal{O}$ (written $\mathcal{I} \models \mathcal{O}$) if $\mathcal{I}$ satisfies every axiom in $\mathcal{O}$. We say that α is a *(logical) consequence* of $\mathcal{O}$ or *is entailed* by $\mathcal{O}$ (written $\mathcal{O} \models \alpha$) if every model of $\mathcal{O}$ satisfies α.

3 Regularity for Sets of Role Inclusion Axioms

Given a set of RIAs $\mathcal{R}$, for every role R, define the following language $L_{\mathcal{R}}(R)$ of role chains (viewed as words over roles):

$$L_{\mathcal{R}}(R) := \{\rho \mid \rho \sqsubseteq^*_{\mathcal{R}} R\} \tag{1}$$

We say that $\mathcal{R}$ is *regular* if the language $L_{\mathcal{R}}(R)$ is regular for every role R. It has been shown [12] that $\prec$-regularity for $\mathcal{R}$ implies regularity for $\bar{\mathcal{R}}$. The converse of this property, however, does not always hold, as demonstrated in the following example.

Consider the following set $\mathcal{R}$ of RIAs:

$$\mathsf{isProperPartOf} \sqsubseteq \mathsf{isPartOf} \tag{2}$$

$$\mathsf{isPartOf} \cdot \mathsf{isPartOf} \sqsubseteq \mathsf{isPartOf} \tag{3}$$

$$\mathsf{isPartOf} \cdot \mathsf{isProperPartOf} \sqsubseteq \mathsf{isProperPartOf} \tag{4}$$

$\mathcal{R}$ expresses properties of parthood relations $\mathsf{isPartOf}$ and $\mathsf{isProperPartOf}$: RIA (2) says that $\mathsf{isProperPartOf}$ is a sub-relation of $\mathsf{isPartOf}$; RIA (3) says that $\mathsf{isPartOf}$ is transitive; RIA (4) says that if x is a part of y which is a proper part of z, then x is a proper part of z. Since any role chain consisting of $\mathsf{isPartOf}$ and $\mathsf{isProperPartOf}$ can be reduced using (2) and (3) to $\mathsf{isPartOf}$, it is easy to see that:

$$L_{\bar{\mathcal{R}}}(\mathsf{isPartOf}) = (\mathsf{isPartOf} \mid \mathsf{isProperPartOf})^+ \tag{5}$$

Since $\mathsf{isProperPartOf}$ is only implied by (4), we also have:

$$L_{\bar{\mathcal{R}}}(\mathsf{isProperPartOf}) = (\mathsf{isPartOf}^* \cdot \mathsf{isProperPartOf})^+ \tag{6}$$

Thus, the languages (5) and (6) induced by RIAs (2)–(4) are regular. However, there is no order $\prec$ for which RIAs (2)–(4) are $\prec$-regular. Indeed, by conditions (i)–(v) of $\prec$-regularity, it follows from (2) that $\mathsf{isProperPartOf} \prec \mathsf{isPartOf}$, and from (4) that $\mathsf{isPartOf} \prec \mathsf{isProperPartOf}$, which is not possible if $\prec$ is a transitive irreflexive relation.

In fact, there is no set of RIAs $\mathcal{R}$, possibly with additional roles, that could express properties (2)–(4) using only $\prec$-regular RIAs. It is easy to show by induction over the definition of $\sqsubseteq^*_{\mathcal{R}}$ that if the RIAs of $\mathcal{R}$ are $\prec$-regular, then $R_1 \cdots R_n \sqsubseteq^*_{\bar{\mathcal{R}}} R$ implies that for every i with $1 \leq i \leq n$, either $R_i = R$, or $R_i = \mathsf{Inv}(R)$, or $R_i \prec R$. This means that for every role R, the language $L_{\bar{\mathcal{R}}}(R)$ contains only words over R, $\mathsf{Inv}(R)$, or R' with $R' \prec R$. Clearly, this is not possible if $L_{\bar{\mathcal{R}}}(\mathsf{isPartOf})$ and $L_{\bar{\mathcal{R}}}(\mathsf{isProperPartOf})$ are extensions of the languages defined in (5) and (6).

Axioms such as (2)–(4) naturally appear in ontologies describing parthood relations, such as those between anatomical parts of the human body. For example, release 7 of the GRAIL version of the OpenGALEN ontology[3] contains the following axioms, which are analogous to (2)–(4):

$$\mathsf{isNonPartitivelyContainedIn} \sqsubseteq \mathsf{isContainedIn} \tag{7}$$

$$\mathsf{isContainedIn} \cdot \mathsf{isContainedIn} \sqsubseteq \mathsf{isContainedIn} \tag{8}$$

$$\mathsf{isNonPartitivelyContainedIn} \cdot \mathsf{isContainedIn} \sqsubseteq \mathsf{isNonPartitivelyContainedIn} \tag{9}$$

[3] `http://www.opengalen.org/`

Complex RIAs such as (7)–(9) are used in OpenGALEN to propagate properties over chains of various parthood relations. For example, the next axiom taken from OpenGALEN expresses that every instance of body structure contained in spinal canal is a structural component of nervous system:

$$\begin{aligned}\mathsf{BodyStructure} \sqcap \exists \mathsf{isContainedIn}.\mathsf{SpinalCanal} \\ \sqsubseteq \exists \mathsf{isStructuralComponentOf}.\mathsf{NervousSystem}\end{aligned} \quad (10)$$

Recently, complex RIAs over parthood relations have been proposed as an alternative to SEP-triplet encoding [19]. The SEP-triplet encoding was introduced [17] as a technique to enable the propagation of some properties over parthood relations, while ensuring that other properties are not propagated. For example, if a finger is defined as part of a hand, then any injury to a finger should be classified as an injury to the hand, however, the amputation of a finger should not be classified as an amputation of the hand. The proposed new encoding makes use of complex RIAs such as (2)–(4) to express propagation properties. For example, propagation of the injury property over the proper-part relation can be expressed using the following RIA:

$$\mathsf{isInjuryOf} \cdot \mathsf{isProperPartOf} \sqsubseteq \mathsf{isInjuryOf}. \quad (11)$$

It was argued that the usage of complex RIAs can eliminate many potential problems with the existing SEP-triplet encoding, used, e.g., in SNOMED CT,[4] and can dramatically reduce the size of the ontology. However, since RIAs (2)–(4) do not satisfy $\prec$-regularity, this technique is currently limited to $\mathcal{EL}^{++}$ ontologies where $\prec$-regularity is not required, and can be problematic when an expressivity beyond $\mathcal{EL}^{++}$ is required, such as for translating OpenGALEN into OWL 2. In this paper we propose an extension of regularity conditions, which, in particular, can handle axioms such as (2)–(4).

4 Stratified Sets of Role Inclusion Axioms and Regularity

As can be seen from example RIAs (2)–(4), one limitation of $\prec$-regularity is that it cannot deal with cyclic dependencies on roles. Our first step is to relax this requirement by considering arbitrary (i.e., not necessarily strict) orders on roles.

Definition 1. *Let $\precsim$ be a preorder (a transitive reflexive relation) on roles. We write $R_1 \approx R_2$ if $R_1 \precsim R_2$ and $R_2 \precsim R_1$, and $R_1 \prec R_2$ if $R_1 \precsim R_2$ and $R_2 \not\precsim R_1$. The* level *$l_{\precsim}(R)$* of *R w.r.t. $\precsim$ is the largest n such that there exists roles $R_1 \prec R_2 \prec \cdots \prec R_n \prec R$. We say that a RIA $R_1 \cdots R_n \sqsubseteq R$ is $\precsim$*-admissible *if $R_i \precsim R$ $(1 \le i \le n)$.*

Unlike $\prec$-regularity, however, $\precsim$-admissibility is not sufficient for regularity since every RIA is $\precsim$-admissible for the *total* preorder $\precsim$, i.e., the one such that $R_1 \precsim R_2$ holds for all roles R_1 and R_2. Note $l_{\precsim}(R) = 0$ for every role R w.r.t. to this preorder. To regain regularity, we impose an additional condition on the set of RIAs $\mathcal{R}$ as a whole.

[4] http://www.ihtsdo.org/

Definition 2. *Given a set of $\precsim$-admissible RIAs $\mathcal{R}$, we say that a RIA $\rho \sqsubseteq R'$ is* $\precsim$-stratified *in $\mathcal{R}$, if for every $R \approx R'$ such that $\rho = \rho_1 R \rho_2$, there exists R_1 such that $\rho_1 R \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1 \rho_2 \sqsubseteq^*_{\mathcal{R}} R'$. We say that $\mathcal{R}$ is* $\precsim$-stratified *if every RIA $\rho \sqsubseteq R$ provable in $\mathcal{R}$, is $\precsim$-stratified in $\mathcal{R}$.*

Intuitively, a set of RIAs $\mathcal{R}$ is $\precsim$-stratified, if every RIA $\rho \sqsubseteq R'$ provable in $\mathcal{R}$ is always provable in $\mathcal{R}$ when reducing the left-most roles of the maximal level first. For example, consider the set $\mathcal{R}$ consisting of RIAs (2)–(4) and (11) and the preorder $\precsim$ such that $\mathsf{isPartOf} \approx \mathsf{isProperPartOf} \precsim \mathsf{isInjuryOf}$ w.r.t. which $\mathcal{R}$ is clearly stratified. Then both of the following RIAs are provable in $\mathcal{R}$ and are $\precsim$-stratified in $\mathcal{R}$:

$$\mathsf{isPartOf} \cdot \mathsf{isPartOf} \cdot \mathsf{isPropertPartOf} \sqsubseteq \mathsf{isProperPartOf}, \tag{12}$$

$$\mathsf{isInjuryOf} \cdot \mathsf{isPartOf} \cdot \mathsf{isPropertPartOf} \sqsubseteq \mathsf{isInjuryOf}. \tag{13}$$

RIA (12) is stratified because for $\rho_1 := \mathsf{isPartOf}$, $R := \mathsf{isPartOf} \approx \mathsf{isProperPartOf} =: R'$, and $\rho_2 := \mathsf{isProperPartOf}$, we have $\rho_1 R = \mathsf{isPartOf} \cdot \mathsf{isPartOf} \sqsubseteq_{(3)} \mathsf{isPartOf} := R_1$ and $R_1 \rho_2 = \mathsf{isPartOf} \cdot \mathsf{isProperParOf} \sqsubseteq_{(4)} \mathsf{isProperPartOf} = R'$. Note that when either $\rho_1 = \epsilon$ or $\rho_2 = \epsilon$, the conditions of Definition 2 hold trivially. RIA (13) is stratified because $R \approx R' := \mathsf{isInjuryOf}$ holds only for $R = \mathsf{isInjuryOf}$, in which case $\rho_1 = \epsilon$. It can be similarly shown that every RIA provable in $\mathcal{R}$, is $\precsim$-stratified in $\mathcal{R}$, so $\mathcal{R}$ is $\precsim$-stratified. If we, however, extend the preorder $\precsim$ such that $\mathsf{isPartOf} \approx \mathsf{isProperPartOf} \approx \mathsf{isInjuryOf}$, i.e., take the total preorder $\precsim$, RIA (13) will be no longer $\precsim$-stratified. Indeed, for $\rho_1 := \mathsf{isInjuryOf}$, $R := \mathsf{isPartOf} \approx \mathsf{isInjuryOf} := R'$, and $\rho_2 := \mathsf{isProperPartOf}$, there does not exist R_1 such that $\rho_1 R = \mathsf{isInjuryOf} \cdot \mathsf{isPartOf} \sqsubseteq^*_{\mathcal{R}} R_1$.

As seen from this example, the choice of the preorder $\precsim$ has an impact on whether $\mathcal{R}$ is $\precsim$-stratified or not. As we pointed out, every RIA is $\precsim$-admissible for the total preorder $\precsim$. However, since all roles R have the same maximal level $L_{\precsim}(R) = 0$ for the total preorder $\precsim$, to check if $\rho \sqsubseteq R'$ is $\precsim$-stratified, one has to consider every role R in ρ, and prove that $\rho_1 R \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1 \rho_2 \sqsubseteq^*_{\mathcal{R}} R'$ hold for the respective prefix ρ_1 and suffix ρ_2. On the other hand, by taking the *smallest* preorder $\precsim_{\mathcal{R}}$ for which the RIAs in $\mathcal{R}$ are $\precsim_{\mathcal{R}}$-admissible, one can avoid many of these tests. The *smallest preorder $\precsim_R$ for $\mathcal{R}$* can be defined as the transitive reflexive closure of the relation $\prec_{\mathcal{R}}$ such that $R_1 \prec_{\mathcal{R}} R_2$ iff $\rho_1 R_1 \rho_2 \sqsubseteq R_2 \in \mathcal{R}$ for some ρ_1 and ρ_2. It can easily be shown using Definition 1 and Definition 2 that for every preorder $\precsim$, (i) all RIAs in $\mathcal{R}$ are $\precsim$-admissible iff $\precsim$ extends $\precsim_{\mathcal{R}}$, and (ii) if $\mathcal{R}$ is $\precsim$-stratified then $\mathcal{R}$ is $\precsim_{\mathcal{R}}$-stratified. From (ii) it follows, in particular, that $\mathcal{R}$ is $\precsim$-stratified for *some* order $\precsim$ iff $\mathcal{R}$ is $\precsim_{\mathcal{R}}$-stratified.

Our next goal is to prove that every $\precsim$-stratified set of RIAs $\mathcal{R}$ induces a regular language $L_{\mathcal{R}}(R)$ for every role R. From now on, we assume that we are given a fixed preorder $\precsim$ and a set of $\precsim$-admissible RIAs $\mathcal{R}$. So, when we say that $\mathcal{R}$ is stratified or a RIA is stratified, we mean $\mathcal{R}$ is $\precsim$-stratified and the RIA is $\precsim$-stratified in $\mathcal{R}$.

First, we distinguish two types of RIAs according to the levels of their roles:

Definition 3. *The* level *of a RIA $\alpha = (R_1 \cdots R_n \sqsubseteq R) \in \mathcal{R}$ (w.r.t. $\precsim$) is $l_{\precsim}(\alpha) := l_{\precsim}(R)$. We say that α is* simple *if $R_i \prec R$ for all i with $1 \leq i \leq (n-1)$; otherwise we say that α is* complex. *For $n \geq 0$, define $\mathcal{R}_n := \{\alpha \in \mathcal{R} \mid l_{\precsim}(\alpha) = n\}$, $\mathcal{R}_{<n} := \bigcup_{k<n} \mathcal{R}_k$, and define $\mathcal{R}^s_n$ to be the set of simple RIAs in $\mathcal{R}_n$.*

In the next lemma, we demonstrate that for every stratified set of RIAs w.l.o.g. one can assume a certain precedence on RIAs in proofs: RIAs of smaller level are applied first; simple RIAs are applied before complex RIAs of the same level; and complex RIAs are only applied to the prefix of the role chain, i.e, at the position 0.

Lemma 1. *For every ρ and R' such that $\rho \sqsubseteq^*_{\mathcal{R}} R'$, there exist ρ^1 and ρ^2 such that $\rho \sqsubseteq^*_{\mathcal{R}_{<n}} \rho^1 \sqsubseteq^*_{\mathcal{R}^s_n} \rho^2 \sqsubseteq^*_{\mathcal{R}_n,0} R'$, where $n = l_{\precsim}(R')$.*

Proof. W.l.o.g., one can assume that $\mathcal{R}$ does not contain RIAs of the form $\epsilon \sqsubseteq R$. Indeed, otherwise for $\mathcal{R}^\epsilon := \{\epsilon \sqsubseteq R \in \mathcal{R}\}$ and $\mathcal{R}' := \mathcal{R} \setminus \mathcal{R}^\epsilon$, we have $\rho \sqsubseteq^*_{\mathcal{R}} R'$ iff $\rho \sqsubseteq^*_{\mathcal{R}^\epsilon} \rho_1 \sqsubseteq^*_{\mathcal{R}'} R'$ for some ρ_1. Now if the lemma holds for $\mathcal{R}'$, then there exist ρ^1_1 and ρ^2_1 such that $\rho_1 \sqsubseteq^*_{\mathcal{R}'_{<n}} \rho^1_1 \sqsubseteq^*_{\mathcal{R}'^s_n} \rho^2_1 \sqsubseteq^*_{\mathcal{R}'_{n,0}} R'$. In this case, it can be readily seen that $\rho \sqsubseteq^*_{\mathcal{R}^\epsilon_{<n}} \rho^1 \sqsubseteq^*_{\mathcal{R}'_{<n}} \rho^2 \sqsubseteq^*_{\mathcal{R}^\epsilon_n} \rho^3 \sqsubseteq^*_{\mathcal{R}'^s_n} \rho^2_1 \sqsubseteq^*_{\mathcal{R}'_{n,0}} R'$ for some ρ^1, ρ^2, and ρ^3.

Now, consider all ρ' such that $\rho \sqsubseteq^*_{\mathcal{R}} \rho' \sqsubseteq^*_{\mathcal{R}} R'$. Since $\mathcal{R}$ does not contain RIAs of the form $\epsilon \sqsubseteq R$, the number of all such ρ' is bounded. From all such ρ', select all $\rho' = \rho'_0 R_1 \rho'_1 \cdots R_m \rho'_m$ with the maximal number of occurrences $R_1, \ldots, R_m$ of roles of level n, and from them select one of the largest length. Then $\rho = \rho_0 \rho^1_1 \rho_1 \cdots \rho^1_m \rho_m$ such that $\rho_i \sqsubseteq^*_{\mathcal{R}_{<n}} \rho'_i$ $(0 \le i \le m)$ and $\rho^1_i \sqsubseteq^*_{\mathcal{R}_{<n}} \rho^2_i \sqsubseteq^*_{\mathcal{R}^s_n} R_i$ $(1 \le i \le m)$. Otherwise one could find ρ' with more occurrences of roles of level n or the same number of occurrences but of a larger length.

Since $\rho'_0 R_1 \rho'_1 \cdots R_m \rho'_m \sqsubseteq^*_{\mathcal{R}} R'$ and $\mathcal{R}$ is stratified, there exist R'_i $(1 \le i \le m)$ such that $\rho'_0 R_1 \sqsubseteq^*_{\mathcal{R}} R'_1$, $R'_i \rho'_i R_{i+1} \sqsubseteq^*_{\mathcal{R}} R'_{i+1}$ $(1 \le i < m)$, and $R'_m \rho'_m \sqsubseteq^*_{\mathcal{R}} R'$. In particular, there exist ρ'^1_i $(1 \le i \le m)$, ρ'^2_i $(0 \le i < m)$ that do not contain roles of level n, and R^1_i $(1 < i \le m)$ such that $\rho'_0 \sqsubseteq^*_{\mathcal{R}_{<n}} \rho'^2_0$, $\rho'^2_0 R_1 \sqsubseteq^*_{\mathcal{R}^s_n} R'_1$, $\rho'_i \sqsubseteq^*_{\mathcal{R}_{<n}} \rho'^1_i \rho'^2_i$, $\rho'^2_i R_{i+1} \sqsubseteq^*_{\mathcal{R}^s_n} R^1_{i+1}$, $R'_i \rho'^1_i R^1_{i+1} \sqsubseteq^*_{\mathcal{R}_n,0} R'_{i+1}$ $(1 \le i < m)$, $\rho'_m \sqsubseteq^*_{\mathcal{R}_{<n}} \rho'^1_m$, and $R'_m \rho'^1_m \sqsubseteq^*_{\mathcal{R}_n,0} R'$. Otherwise one could again find ρ' with more roles of level n or with the same number of roles but of a larger length.

Summing up, we obtain the required ρ^1 and ρ^2 as follows:

$$\begin{aligned}\rho =& \rho_0 \rho^1_1 \rho_1 \cdots \rho^1_m \rho_m \sqsubseteq^*_{\mathcal{R}_{<n}} \rho^1 := \rho'^2_0 \rho^2_1 \rho'^1_1 \rho'^2_1 \cdots \rho^2_m \rho'^1_m \\ & \sqsubseteq^*_{\mathcal{R}^s_n} \rho^2 := R'_1 \rho'^1_1 R^1_2 \rho'^1_2 \cdots R^1_m \rho'^1_m \sqsubseteq^*_{\mathcal{R}_n,0} R'.\end{aligned}$$ □

We are now in a position to prove that every stratified set of RIAs $\mathcal{R}$ is regular.

Theorem 1. *For every $\precsim$-stratified set of RIAs $\mathcal{R}$ and every role R, one can construct a non-deterministic finite automaton (NFA) that recognizes the language $L_{\mathcal{R}}(R)$. The size (i.e., the number of transitions) of the automaton is bounded by $(c \cdot m)^{2 \cdot n}$ where $m := \|\mathcal{R}\|$, $n := l_{\precsim}(R)$, and c is some fixed constant.*

Proof. By Lemma 1, for every role R we have:

$$L_{\mathcal{R}}(R) = \{\rho \mid \exists \rho^1 \exists \rho^2 : \rho \sqsubseteq^*_{\mathcal{R}_{<n}} \rho^1 \sqsubseteq^*_{\mathcal{R}^s_n} \rho^2 \sqsubseteq^*_{\mathcal{R}_n,0} R\}. \quad (14)$$

We first show that the languages $L_{\mathcal{R}^s_n}(R)$ and $L_{\mathcal{R}_n,0}(R) := \{\rho \mid \rho \sqsubseteq^*_{\mathcal{R}_n,0} R\}$ can be recognized by NFAs. For every role R, introduce a terminal symbol a_R and a non-terminal symbol A_R. For $\mathcal{R}^s_n$, consider a grammar containing production rules $A_R \to a_R$

for every role R and $A_R \to a_{R_1} \cdots a_{R_{k-1}} A_{R_k}$ for every RIA $R_1 \cdots R_k \sqsubseteq R \in \mathcal{R}_n^s$. Since for every simple RIA $R_1 \cdots R_k \sqsubseteq R \in \mathcal{R}_n^s$ the level of the roles R_i with $1 \leq i < k$ is smaller than n, it is easy to show by induction that $R_1 \cdots R_m \in L_{\mathcal{R}_n^s}(R)$ iff $A_R \to^* a_{R_1} \ldots a_{R_m}$. Since the grammar is right-linear, the language $L_{\mathcal{R}_n^s}(R)$ is regular. Similarly, the language $L_{\mathcal{R}_n,0}(R)$ is regular since it corresponds to the left-linear grammar containing production rules $A_R \to a_R$ for every role R and $A_R \to A_{R_1} a_{R_2} \cdots a_{R_k}$ for every RIA $R_1 \cdots R_k \sqsubseteq R \in \mathcal{R}_n$. It is well-known that for left- and right-linear grammars one can construct an NFA with size linear in the size of the grammar, which in our cases is bounded by $c \cdot m$ where $m := \|\mathcal{R}\|$ and c a constant.

Assume, by induction hypothesis, that for every R such that $l_{\precsim}(R) < n$, $L_{\mathcal{R}}(R)$ can be recognized by an NFA of size at most $(c \cdot m)^{2 \cdot (n-1)}$. Since $\widetilde{L}_{\mathcal{R}_{<n}}(R) = \{R\}$ if $l_{\precsim}(R) \geq n$ and $L_{\mathcal{R}_{<n}}(R) = L_{\mathcal{R}}(R)$ if $l_{\precsim}(R) < n$, for $L_{\mathcal{R}_{<n}}(R)$ one can construct an NFA of size bounded by $(c \cdot m)^{2 \cdot (n-1)}$. For the remaining case $l_{\precsim}(R) = n$, consider $L_n(R) := \{\rho^1 \mid \exists \rho^2 : \rho^1 \sqsubseteq^*_{\mathcal{R}_n^s} \rho^2 \sqsubseteq^*_{\mathcal{R}_n,0} R\}$. Clearly, $\rho^1 \in L_n(R)$ iff there exists $\rho^2 = R_1 \cdots R_k \in L_{\mathcal{R}_n,0}(R)$ such that $\rho^1 = \rho^1_1 \cdots \rho^1_k$ where $\rho^1_i \in L_{\mathcal{R}_n^s}(R_i)$ $(1 \leq i \leq k)$. Hence $L_n(R) = L_{\mathcal{R}_n,0}(R)[R_1/L_{\mathcal{R}_n^s}(R_1), \ldots, R_k/L_{\mathcal{R}_n^s}(R_k)]$ where R_i are all roles of level n and $L[R_1/L_1, \ldots, R_k/L_k]$ denotes the language obtained by substituting in every word from L the letters R_i with words from L_i in all possible ways. Since regular languages are closed under substitution and an NFA for $L[R_1/L_1, \ldots, R_m/L_m]$ can be constructed with the size bounded by the size of the NFA for L multiplied with the maximum size of the NFA for $L_1, \ldots, L_m$, the language $L_n(R)$ is regular and can be recognized by an NFA of size at most $(c \cdot m)^2$.

Similarly, by (14), we have $L_{\mathcal{R}}(R) = L_n(R)[R_1/L_{\mathcal{R}_{<n}}(R_1), \ldots, R_k/L_{\mathcal{R}_{<n}}(R_k)]$ where R_i for $1 \leq i \leq k$ are all roles of level n. Thus, one can construct an NFA of size bounded by $(c \cdot m)^2 \cdot (c \cdot m)^{2 \cdot (n-1)} = (c \cdot m)^{2 \cdot n}$ that recognizes $L_{\mathcal{R}}(R)$. □

5 Testing if a Set of RIAs is Stratified

Up to now we have demonstrated that, similar to $\prec$-regularity, Definition 2 provides a sufficient condition for regularity of a set of RIAs $\mathcal{R}$. However, unlike $\prec$-regularity, Definition 2 does not provide for any effective means of testing this condition since it requires to test if all (of possibly infinitely many) RIAs $\rho \sqsubseteq R$ provable in $\mathcal{R}$, are stratified. Below we demonstrate that it suffices to test regularity only for finitely many RIAs that can be effectively computed from $\mathcal{R}$.

Definition 4. *Let $\mathcal{R}$ be a set of $\precsim$-admissible RIAs. RIA $\rho_1 R \rho_2 \sqsubseteq R'$ is an* overlap *of two RIAs $\rho_2^\epsilon R \rho_2 \sqsubseteq R_1$ and $\rho_1 R_2 \rho_1^\epsilon \sqsubseteq R'$ (w.r.t. $\mathcal{R}$ and $\precsim$) if $R \approx R'$, $\epsilon \sqsubseteq^*_{\mathcal{R}} \rho_2^\epsilon$, $\epsilon \sqsubseteq^*_{\mathcal{R}} \rho_1^\epsilon$, and $R_1 \sqsubseteq^*_{\mathcal{R}} R_2$. $\mathcal{R}$ is* weakly $\precsim$-stratified *if (i) every RIA in $\mathcal{R}$ is $\precsim$-stratified in $\mathcal{R}$ and (ii) the overlap of every two RIAs in $\mathcal{R}$ is $\precsim$-stratified in $\mathcal{R}$.*

Note that the overlap $\rho_1 R \rho_2 \sqsubseteq R'$ is provable in $\mathcal{R}$ in a such a way that only RIAs $\rho_2^\epsilon R \rho_2 \sqsubseteq R_1$ and $\rho_1 R_2 \rho_1^\epsilon \sqsubseteq R'$ involved in the overlap reduce the length of the role chains: $\rho_1(R\rho_2) \sqsubseteq^*_{\mathcal{R}} \rho_1(\rho_2^\epsilon R \rho_2) \sqsubseteq_{\mathcal{R}} \rho_1 R_1 \sqsubseteq^*_{\mathcal{R}} \rho_1 R_2 \sqsubseteq^*_{\mathcal{R}} \rho_1 R_2 \rho_1^\epsilon \sqsubseteq_{\mathcal{R}} R'$. Intuitively, to prove that $\mathcal{R}$ is weakly stratified, one has to consider only RIAs provable in $\mathcal{R}$ using at most two reducing steps, first, reducing the suffix of a role chain, and second, reducing the prefix of the role chain: $\rho_1(R\rho_2) \sqsubseteq^*_{\mathcal{R}} \rho_1 R_1 \sqsubseteq^*_{\mathcal{R}} R'$.

For example, the set $\mathcal{R}$ of RIAs (2)–(4) is weakly stratified. Indeed, every RIA in $\mathcal{R}$ is trivially stratified since no role chain in $\mathcal{R}$ has more than two roles. RIAs (3) and (4) can overlap (possibly with themselves) only in the following three cases:

$$\mathsf{isPartOf} \cdot (\mathsf{isPartOf} \cdot \mathsf{isPartOf}) \sqsubseteq_{(2)} \mathsf{isPartOf} \cdot \mathsf{isPartOf} \sqsubseteq_{(3)} \mathsf{isPartOf}, \tag{15}$$

$$\begin{aligned} &\mathsf{isPartOf} \cdot (\mathsf{isPartOf} \cdot \mathsf{isProperPartOf}) \sqsubseteq_{(4)} \\ &\qquad \mathsf{isPartOf} \cdot \mathsf{isProperPartOf} \sqsubseteq_{(4)} \mathsf{isProperPartOf}, \end{aligned} \tag{16}$$

$$\begin{aligned} &\mathsf{isPartOf} \cdot (\mathsf{isPartOf} \cdot \mathsf{isProperPartOf}) \sqsubseteq_{(4)} \\ &\qquad \mathsf{isPartOf} \cdot \mathsf{isProperPartOf} \sqsubseteq_{(2)} \\ &\qquad\qquad \mathsf{isPartOf} \cdot \mathsf{isPartOf} \sqsubseteq_{(3)} \mathsf{isPartOf}. \end{aligned} \tag{17}$$

The resulted overlaps are provable using (2)–(4) "left-to-right" and thus stratified:

$$(\mathsf{isPartOf} \cdot \mathsf{isPartOf}) \cdot \mathsf{isPartOf} \sqsubseteq_{(3),(3)} \mathsf{isPartOf}, \tag{18}$$

$$(\mathsf{isPartOf} \cdot \mathsf{isPartOf}) \cdot \mathsf{isProperPartOf} \sqsubseteq_{(3),(4)} \mathsf{isProperPartOf}, \tag{19}$$

$$(\mathsf{isPartOf} \cdot \mathsf{isPartOf}) \cdot \mathsf{isProperPartOf} \sqsubseteq_{(3),(4),(2)} \mathsf{isPartOf}. \tag{20}$$

Similarly, one can show that $\bar{\mathcal{R}}$ is weakly stratified by considering all overlaps between the inverses of (3) and (4).

The notion of overlap and conditions for a weakly stratified set of RIAs are reminiscent of the well-known notions of a *critical pair* and the *weak Church-Rosser property* from term rewriting [3]. Despite close resemblance, there seem, however, to be no direct correspondence between these properties—if to consider the entailment relation $\sqsubseteq_{\mathcal{R}}$ as a rewriting relation on chains of roles, the conditions of Definition 4 essentially mean that if a role chain can be rewritten to a role using at most two complex "rightmost" reductions, then it can also be rewritten to the same role chain using only "leftmost" reductions. Like in Church-Rosser theorem, however, it is possible to prove that every weakly stratified set of RIAs is stratified:

Theorem 2. *For every preorder $\precsim$ and every set of $\precsim$-admissible RIAs $\mathcal{R}$, $\mathcal{R}$ is $\precsim$-stratified iff $\mathcal{R}$ is weakly $\precsim$-stratified.*

Proof. As in the proof of Lemma 1, we first prove that w.l.o.g. one can assume that $\mathcal{R}$ does not contain RIAs of the form $\epsilon \sqsubseteq R$. Otherwise, we extend $\mathcal{R}$ by repeatedly adding for every $\epsilon \sqsubseteq R \in \mathcal{R}$ and $\rho_1 R \rho_2 \sqsubseteq R' \in \mathcal{R}$ a RIA $\rho_1\rho_2 \sqsubseteq R$. This transformation preserves the set of implied RIAs, and therefore the result $\mathcal{R}'$ is stratified iff $\mathcal{R}$ is stratified. It can also be shown that if $\rho \sqsubseteq R'$ is provable in $\mathcal{R}'$ and $\rho \neq \epsilon$ then $\rho \sqsubseteq R'$ is provable in $\mathcal{R}'$ without using RIAs of the form $\epsilon \sqsubseteq R$. Hence, if we prove that is $\mathcal{R}'$ weakly stratified iff $\mathcal{R}$ is weakly stratified, we can disregard all axioms of the form $\epsilon \sqsubseteq R$ in this proof (since all provable RIAs $\rho \sqsubseteq R$ for $\rho = \epsilon$ are trivially stratified).

It is clear that if $\mathcal{R}'$ is weakly stratified then $\mathcal{R}$ is as well since $\mathcal{R}'$ contains $\mathcal{R}$. To prove the converse, assume that $\mathcal{R}$ is weakly stratified. Let $\mathcal{R}^\epsilon := \{\epsilon \sqsubseteq R \in \mathcal{R}\}$. Note that if $\rho' \sqsubseteq_{R^\epsilon} \rho$ and $\rho \sqsubseteq R$ is stratified then $\rho' \sqsubseteq R$ is stratified as well (in both $\mathcal{R}$ and $\mathcal{R}'$). Hence the condition (i) of Definition 4 for $\mathcal{R}'$ is immediate from this property and the construction of $\mathcal{R}'$. To prove condition (ii) of Definition 4 for $\mathcal{R}'$, let $\rho_1 R \rho_2 \sqsubseteq R'$

be an overlap of two RIAs $\rho_2^\epsilon R\rho_2 \sqsubseteq R_1$ and $\rho_1 R_2 \rho_1^\epsilon \sqsubseteq R'$ in $\mathcal{R}'$. From the construction of $\mathcal{R}'$, there should be RIAs $\rho_4^\epsilon R\rho_4 \sqsubseteq R_1$ and $\rho_3 R_2 \rho_3^\epsilon \sqsubseteq R'$ in $\mathcal{R}$ such that $\rho_2^\epsilon \sqsubseteq^*_{\mathcal{R}^\epsilon} \rho_4^\epsilon$, $\rho_2 \sqsubseteq^*_{\mathcal{R}^\epsilon} \rho_4$, $\rho_1 \sqsubseteq^*_{\mathcal{R}^\epsilon} \rho_3$, and $\rho_1^\epsilon \sqsubseteq^*_{\mathcal{R}^\epsilon} \rho_3^\epsilon$. In particular $\epsilon \sqsubseteq^*_{\mathcal{R}} \rho_2^\epsilon \sqsubseteq^*_{\mathcal{R}} \rho_4^\epsilon$, $\epsilon \sqsubseteq^*_{\mathcal{R}} \rho_1^\epsilon \sqsubseteq^*_{\mathcal{R}} \rho_3^\epsilon$, and so, $\rho_3 R\rho_4 \sqsubseteq R'$ is an overlap of RIAs in $\mathcal{R}$, which by condition (ii) should be stratified. Since $\rho_1 R\rho_2 \sqsubseteq^*_{\mathcal{R}^\epsilon} \rho_3 R\rho_4$, we obtain that $\rho_1 R\rho_2 \sqsubseteq R'$ is stratified as well.

So now, w.l.o.g., we can assume that $\mathcal{R}$ does not contain RIAs of the form $\epsilon \sqsubseteq R$.

The "only if" direction of the theorem is trivial since RIAs in $\mathcal{R}$ and overlaps of RIAs in $\mathcal{R}$ are provable in $\mathcal{R}$.

To prove the "if" direction, assume to the contrary that there exists ρ such that $\rho \sqsubseteq^*_{\mathcal{R}} R'$ but $\rho \sqsubseteq R'$ is not stratified w.r.t. $\mathcal{R}$ and $\precsim$.

Take such a ρ of the smallest length. Then $\rho = \rho_1 R\rho_2$ where $R \approx R'$ and there exists no R_1 such that $\rho_1 R \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1\rho_2 \sqsubseteq^*_{\mathcal{R}} R'$. Clearly, $\rho_1 \neq \epsilon$ and $\rho_2 \neq \epsilon$.

Since $\rho = \rho_1 R\rho_2 \sqsubseteq^*_{\mathcal{R}} R'$, there exist ρ_1^1, ρ_1^2, ρ_2^1, ρ_2^2, R^1, and R^2 such that $\rho_1 \sqsubseteq^*_{\mathcal{R}} \rho_1^2\rho_1^1$, $\rho_2 \sqsubseteq^*_{\mathcal{R}} \rho_2^1\rho_2^2$, $R \sqsubseteq^*_{\mathcal{R}} R^1$, $\rho_1^1 R^1 \rho_2^1 \sqsubseteq R^2 \in \mathcal{R}$, $\rho_1^2 R^2 \rho_2^2 \sqsubseteq^*_{\mathcal{R}} R'$, and $\rho_1^1\rho_2^1 \neq \epsilon$, so:

$$\rho_1 R\rho_2 \sqsubseteq^*_{\mathcal{R}} \rho_1^2(\rho_1^1 R^1 \rho_2^1)\rho_2^2 \sqsubseteq_{\mathcal{R}} \rho_1^2 R^2 \rho_2^2 \sqsubseteq^*_{\mathcal{R}} R'.$$

Since $\mathcal{R}$ is weakly stratified and $\rho_1^1 R^1 \rho_2^1 \sqsubseteq R^2 \in \mathcal{R}$, by condition (i) in Definition 4, $\rho_1^1 R^1 \sqsubseteq^*_{\mathcal{R}} R_1^1$ and $R_1^1 \rho_2^1 \sqsubseteq^*_{\mathcal{R}} R^2$ for some R_1^1. In particular, $\rho_1^2 R_1^1 \rho_2^1 \rho_2^2 \sqsubseteq^*_{\mathcal{R}} R'$.

We prove that $\rho_1^1 = \epsilon$. If $\rho_1^1 \neq \epsilon$, then $\|\rho_1^2 R_1^1 \rho_2^1 \rho_2^2\| < \|\rho\|$, so, $\rho_1^2 R_1^1 \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1 \rho_2^1 \rho_2^2 \sqsubseteq^*_{\mathcal{R}} R'$ for some R_1. We obtain a contradiction since $\rho_1 R \sqsubseteq^*_{\mathcal{R}} \rho_1^2(\rho_1^1 R^1) \sqsubseteq^*_{\mathcal{R}} \rho_1^2 R_1^1 \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1\rho_2 \sqsubseteq^*_{\mathcal{R}} R_1 \rho_2^1 \rho_2^2 \sqsubseteq^*_{\mathcal{R}} R'$, but we assumed that no such R_1 exists. Thus $\rho_1^1 = \epsilon$.

Now we prove that $\rho_2^2 = \epsilon$. Since $\rho_1^1\rho_2^1 \neq \epsilon$, $\|\rho_1^2 R^2 \rho_2^2\| < \|\rho\|$, so, $\rho_1^2 R^2 \sqsubseteq^*_{\mathcal{R}} R_1^2$ and $R_1^2 \rho_2^2 \sqsubseteq^*_{\mathcal{R}} R'$ for some R_1^2. In particular, $\rho_1^2\rho_1^1 R^1 \rho_2^1 \sqsubseteq^*_{\mathcal{R}} R_1^2$. If $\rho_2^2 \neq \epsilon$ then $\|\rho_1^2\rho_1^1 R^1 \rho_2^1\| < \|\rho\|$, so $\rho_1^2\rho_1^1 R^1 \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1\rho_2^1 \sqsubseteq^*_{\mathcal{R}} R_1^2$ for some R_1. We obtain a contradiction since $\rho_1 R \sqsubseteq^*_{\mathcal{R}} \rho_1^2\rho_1^1 R^1 \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1\rho_2 \sqsubseteq^*_{\mathcal{R}} (R_1\rho_2^1)\rho_2^2 \sqsubseteq^*_{\mathcal{R}} R_1^2\rho_2^2 \sqsubseteq^*_{\mathcal{R}} R'$, but we assumed that no such R_1 exists. Thus $\rho_2^2 = \epsilon$.

Now since $\rho_1^2 R^2 \sqsubseteq^*_{\mathcal{R}} R'$, there exist ρ_1^3, ρ_1^4, R^3, and R^4 such that $\rho_1^2 \sqsubseteq^*_{\mathcal{R}} \rho_1^4\rho_1^3$, $R^2 \sqsubseteq^*_{\mathcal{R}} R^3$, $\rho_1^3 R^3 \sqsubseteq R^4 \in \mathcal{R}$, $\rho_1^4 R^4 \sqsubseteq^*_{\mathcal{R}} R'$, and $\rho_1^3 \neq \epsilon$. So the full picture is:

$$\begin{aligned}\rho_1 R\rho_2 \sqsubseteq^*_{\mathcal{R}} \rho_1^2 R^1 \rho_2^1 \sqsubseteq^*_{\mathcal{R}} \rho_1^4\rho_1^3(R^1\rho_2^1) \sqsubseteq_{\mathcal{R}}& \\ \rho_1^4\rho_1^3 R^2 \sqsubseteq^*_{\mathcal{R}} \rho_1^4(\rho_1^3 R^3) \sqsubseteq_{\mathcal{R}} \rho_1^4 R^4 \sqsubseteq^*_{\mathcal{R}} R'.&\end{aligned}$$

In this case, $\rho_1^3 R^1 \rho_2^1 \sqsubseteq R^4$ is an overlap of the RIAs $R^1\rho_2^1 \sqsubseteq R^2$ and $\rho_1^3 R^3 \sqsubseteq R^4$ w.r.t. $\mathcal{R}$ and $\precsim$. Since $\mathcal{R}$ is weakly stratified, by condition (ii) in Definition 4, $\rho_1^3 R^1 \sqsubseteq^*_{\mathcal{R}} R_1^3$ and $R_1^3\rho_2^1 \sqsubseteq R^4$ for some R_1^3. In particular, $\rho_1^4 R_1^3 \rho_2^1 \sqsubseteq^*_{\mathcal{R}} R'$. Since $p_1^3 \neq \epsilon$, $\|\rho_1^4 R_1^1 \rho_2^1\| < \|\rho\|$, so, $\rho_1^4 R_1^3 \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1\rho_2^1 \sqsubseteq^*_{\mathcal{R}} R'$ for some R_1. We obtain a contradiction since $\rho_1 R \sqsubseteq^*_{\mathcal{R}} \rho_1^4(\rho_1^3 R^1) \sqsubseteq^*_{\mathcal{R}} \rho_1^4 R_1^3 \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1\rho_2 \sqsubseteq^*_{\mathcal{R}} R_1\rho_2^1 \sqsubseteq^*_{\mathcal{R}} R'$, but we assumed that no such R_1 exists. □

Now, in order to present an algorithm for deciding whether a set of RIAs $\mathcal{R}$ is stratified, according to Theorem 2, it is sufficient to prove that one can effectively check the conditions of Definition 4.

Lemma 2. *Given a set of RIAs $\mathcal{R}$ and a RIA $\rho \sqsubseteq R$, it is possible to decide in polynomial time whether $\rho \sqsubseteq^*_{\mathcal{R}} R$.*

Proof. Define a context-free grammar with terminal symbols a_R and non-terminal symbols A_R for every role R, and production rules $A_R \to a_R$ for every role R and $A_R \to A_{R_1} \dots A_{R_n}$ for every RIA $R_1 \cdots R_n \sqsubseteq R \in \mathcal{R}$. It is easy to show that $A_R \to^* a_{R_1} \dots a_{R_n}$ w.r.t. this grammar iff $R_1 \cdots R_n \sqsubseteq^*_{\mathcal{R}} R$. Since the word problem (membership in the language) for context-free grammars is decidable in polynomial time (see, e.g. [10]), so is the property $\rho \sqsubseteq^*_{\mathcal{R}} R$. □

Corollary 1. *For every $\precsim$-admissible set of RIAs $\mathcal{R}$, one can check in polynomial time in $\|\mathcal{R}\|$ if $\mathcal{R}$ is $\precsim$-stratified.*

Proof. By Theorem 2, to check if $\mathcal{R}$ is stratified, it is sufficient to check if every RIA in $\mathcal{R}$ is stratified and every overlap of two RIAs is stratified. Hence there are only polynomially-many RIAs to test. In order to test whether $\rho_1 R \rho_2 \sqsubseteq R'$ is stratified for R, we enumerate all roles R_1 in $\mathcal{R}$ and check if $\rho_1 R \sqsubseteq^*_{\mathcal{R}} R_1$ and $R_1 \rho_2 \sqsubseteq^*_{\mathcal{R}} R'$ hold. By Lemma 2, each of these conditions can be checked in polynomial time. □

Using the criterion in Theorem 2, it is now possible to show that for every set $\mathcal{R}$ of $\prec$-regular RIAs (according to the original conditions in $\mathcal{SROIQ}$), $\bar{\mathcal{R}}$ is stratified w.r.t. $\precsim$ defined by $R_1 \precsim R_2$ if either $R_1 \prec R_2$, $R_1 = R_2$ or $R_1 = \mathsf{Inv}(R_2)$. Clearly, the conditions of $\prec$-regularity ensure that every $\rho \sqsubseteq R \in \bar{\mathcal{R}}$ is stratified. Note also that if $\rho R \sqsubseteq R' \in \bar{\mathcal{R}}$ or $R \rho \sqsubseteq R' \in \bar{\mathcal{R}}$ with $R \approx R'$ then either $R' = R$ or $\rho = \epsilon$. Now, if $\rho_1 R \rho_2 \sqsubseteq R'$ is an overlap of two RIAs $\rho_2^\epsilon R \rho_2 \sqsubseteq R_2$ and $\rho_1 R_1 \rho_1^\epsilon \sqsubseteq R'$ in $\bar{\mathcal{R}}$ with $\rho_1 \neq \epsilon$ and $\rho_2 \neq \epsilon$ (otherwise it is trivially stratified), then $R_2 \sqsubseteq^*_{\bar{\mathcal{R}}} R_1$, $R \approx R_2 \approx R_1 \approx R'$, so $\rho_1^\epsilon = \rho_2^\epsilon = \epsilon$, $R_2 = R$, $R' = R_1$, and either $R_1 = R_2$ or $R_1 = \mathsf{Inv}(R_2)$ (by definition of $\precsim$). From the last and the fact that $R_2 \sqsubseteq^*_{\bar{\mathcal{R}}} R_1$, it follows that $R_1 \sqsubseteq^*_{\bar{\mathcal{R}}} R_2$. Hence, $\rho_1 R = \rho_1 R_2 \sqsubseteq^*_{\bar{\mathcal{R}}} \rho_1 R_1 \sqsubseteq_{\bar{\mathcal{R}}} R' = R_1 \sqsubseteq^*_{\bar{\mathcal{R}}} R_2 = R$, $R \rho_2 \sqsubseteq_{\bar{\mathcal{R}}} R_2 \sqsubseteq^*_{\bar{\mathcal{R}}} R_1 = R'$, and so $\rho_1 R \rho_2 \sqsubseteq R'$ is stratified.

To illustrate the practical benefits of Definition 4, consider the set $\mathcal{R}$ of RIAs (2)–(4) and (11) for the total preorder $\precsim$. As we showed before, $\bar{\mathcal{R}}$ is not stratified because of RIA (13) provable in $\bar{\mathcal{R}}$. RIA (13) can be obtained as an overlap of RIAs (4) and (11):

$$\begin{aligned} &\mathsf{isInjuryOf} \cdot (\mathsf{isPartOf} \cdot \mathsf{isProperPartOf}) \sqsubseteq_{(4)} \\ &\qquad \mathsf{isInjuryOf} \cdot \mathsf{isProperPartOf} \sqsubseteq_{(11)} \mathsf{isInjuryOf}. \end{aligned} \tag{21}$$

RIA (13) is not stratified, because there is no R_1 such that $\mathsf{isInjuryOf} \cdot \mathsf{isPartOf} \sqsubseteq^*_{\bar{\mathcal{R}}} R_1$ and $R_1 \cdot \mathsf{isProperPartPf} \sqsubseteq^*_{\bar{\mathcal{R}}} \mathsf{isInjuryOf}$ hold. In our situation, the roles isInjuryOf and isPartOf can be "composed" together with the third role isProperPartOf, but cannot be composed directly. This typically indicates on some missing properties, which the domain expert, presented with this situation, could often easily identify. In our case, the set of RIAs becomes stratified as soon as we add the axiom propagating the injury relation over the part-of relation, which then subsumes (11) given (2):

$$\mathsf{isInjuryOf} \cdot \mathsf{isPartOf} \sqsubseteq \mathsf{isInjuryOf}. \tag{22}$$

Thus, Definition 4 has two practical benefits. First, it can be used to check automatically if the given set of RIAs is stratified. Second, in the case when the set of RIAs is not stratified, it is possible to use this definition in interactive setting when the user is presented with problematic overlaps and prompted to enter the missing RIAs.

It is a natural question, whether for any set of RIAs $\mathcal{R}$ that induces regular languages, there exists an extension $\mathcal{R}'$, as in this example, that is stratified. The following theorem gives a surprising positive answer to this question. It turns out, there always exists a stratified *conservative extension* of $\mathcal{R}$—a super-set $\mathcal{R}'$ of $\mathcal{R}$ possibly containing new roles, such that any model $\mathcal{I}$ of $\mathcal{R}$ can be extended to a model $\mathcal{J}$ of $\mathcal{R}'$ that interprets the roles occurring in $\mathcal{R}$ exactly as $\mathcal{I}$ does.

Theorem 3. *Let $\mathcal{R}$ be a set of RIAs such that $L_{\bar{\mathcal{R}}}(R)$ is regular for every role R. Then there exists a conservative extension $\mathcal{R}'$ of $\mathcal{R}$ such that $\bar{\mathcal{R}}'$ is $\precsim$-stratified for every $\precsim$.*

Proof. Let Σ be the set of all roles occurring in $\bar{\mathcal{R}}$. For every $R \in \Sigma$ and $\rho_1, \rho_2 \in \Sigma^*$, define the language $L_{\bar{\mathcal{R}}}(R, \rho_1, \rho_2) := \{\rho \mid \rho_1\rho\rho_2 \in L_{\bar{\mathcal{R}}}(R)\}$. It follows from Myhill-Nerode theorem (see, e.g., [18]) that $L_{\bar{\mathcal{R}}}(R)$ is regular iff there are only finitely many different languages $L_{\bar{\mathcal{R}}}(R, \rho_1, \rho_2)$ for all possible ρ_1 and ρ_2. Let $L_{\bar{\mathcal{R}}} := \{L_{\bar{\mathcal{R}}}(R, \rho_1, \rho_2) \mid R \in \Sigma, \rho_1, \rho_2 \in \Sigma^*\}$ be the set of all languages of this form, and $S_{\mathcal{R}} := \{\bigcap_{L \in S} L \mid S \subseteq L_{\bar{\mathcal{R}}}\}$ be the set of all their possible intersections (the empty intersection is Σ^*). Since every language $L_{\bar{\mathcal{R}}}(R)$ is regular, clearly, both sets $L_{\bar{\mathcal{R}}}$ and $S_{\mathcal{R}}$ are finite. Note that $L_{\bar{\mathcal{R}}}(R) = L_{\bar{\mathcal{R}}}(R, \epsilon, \epsilon) \in L_{\bar{\mathcal{R}}} \subseteq S_{\mathcal{R}}$. For languages $L_1, \ldots, L_n \subseteq \Sigma^*$ (not necessarily in $S_{\mathcal{R}}$), let $L_1 \cdots L_n := \{\rho_1 \cdots \rho_n \mid \rho_i \in L_i, 1 \leq i \leq n\}$ if $n > 0$, or $\{\epsilon\}$ if $n = 0$. One nice feature of $S_{\mathcal{R}}$ is the following interpolation-like property:

Claim 1 *For every L_1, L_2, L such that $L_1 \cdot L_2 \subseteq L \in S_{\mathcal{R}}$, there exists $L'_1 \in S_{\mathcal{R}}$ and $L'_2 \in S_{\mathcal{R}}$ such that $L_i \subseteq L'_i$, $i = 1, 2$, and $L'_1 \cdot L'_2 \subseteq L$.*

Indeed, let $L = \bigcap_{1 \leq i \leq n} L_{\bar{\mathcal{R}}}(R_i, \rho_i^1, \rho_i^2)$ for some R_i, ρ_i^1, and ρ_i^2, $(1 \leq i \leq n)$. Define:

$$L'_1 := \bigcap_{\rho_2 \in L_2, 1 \leq i \leq n} L_{\bar{\mathcal{R}}}(R_i, \rho_i^1, \rho_2\rho_i^2); \qquad L'_2 := \bigcap_{\rho_1 \in L'_1, 1 \leq i \leq n} L_{\bar{\mathcal{R}}}(R_i, \rho_i^1\rho_1, \rho_i^2).$$

To prove that $L_1 \subseteq L'_1$, take any $\rho_1 \in L_1$. Since $L_1 \cdot L_2 \subseteq L$, for every $\rho_2 \in L_2$, we have $\rho_1\rho_2 \in L$. By definition of L and $L_{\bar{\mathcal{R}}}(R_i, \rho_i^1, \rho_i^2)$, we have $\rho_i^1\rho_1\rho_2\rho_i^2 \in L_{\bar{\mathcal{R}}}(R_i)$, $(1 \leq i \leq n)$, so $\rho_1 \in L_{\bar{\mathcal{R}}}(R_i, \rho_i^1, \rho_2\rho_i^2)$, $(1 \leq i \leq n)$. Thus $\rho_1 \in L'_1$.

We now prove that $L'_1 \cdot L_2 \subseteq L$. Take any any $\rho_1 \in L'_1$ and $\rho_2 \in L_2$. By definition of L'_1, $\rho_1 \in L_{\bar{\mathcal{R}}}(R_i, \rho_i^1, \rho_2\rho_i^2)$, so $\rho_1\rho_2 \in L_{\bar{\mathcal{R}}}(R_i, \rho_i^1, \rho_i^2)$, $(1 \leq i \leq n)$. Thus $\rho_1\rho_2 \in L$.

Using $L'_1 \cdot L_2 \subseteq L$, it is now easy to show that $L_2 \subseteq L'_2$ and $L'_1 \cdot L'_2 \subseteq L$ symmetrically to the proofs above.

To continue the proof of the theorem, for every $L \in S_{\mathcal{R}}$, we introduce a fresh role R_L. Consider the set $\mathcal{R}_1$ consisting of the following RIAs for every $L, L_1, L_2 \in S_{\mathcal{R}}$:

$$\epsilon \sqsubseteq R_L \qquad \text{if } \epsilon \in L, \tag{23}$$

$$R_{L_1} \sqsubseteq R_L \qquad \text{if } L_1 \subseteq L, \tag{24}$$

$$R_{L_1} \cdot R_{L_2} \sqsubseteq R_L \qquad \text{if } L_1 \cdot L_2 \subseteq L. \tag{25}$$

Claim 2 *For every $L_1, \ldots, L_n, L \in S_{\mathcal{R}}$, $L_1 \cdots L_n \subseteq L$ iff $R_{L_1} \cdots R_{L_n} \sqsubseteq^*_{\mathcal{R}_1} R_L$.*

The "if" direction of the claim can easily be shown by induction on the length of the proof of $R_{L_1} \cdots R_{L_n} \sqsubseteq^*_{\mathcal{R}_1} R_L$.

We prove the "only if" direction of the claim by induction on n. For $n \leq 2$ the claim follows directly from (23)–(25).

Now, if $L_1 \cdots L_n \cdot L_{n+1} \subseteq L$ then by Claim 1, there exists $L' \in S_{\mathcal{R}}$ such that $L_1 \cdots L_n \subseteq L'$ and $L' \cdot L_{n+1} \subseteq L$. By induction hypothesis, $R_{L_1} \cdots R_{L_n} \sqsubseteq^*_{\mathcal{R}_1} R_{L'}$. Since by (25) we have $R_{L'} \cdot R_{L_{n+1}} \sqsubseteq R_L \in \mathcal{R}_1$, we obtain that $R_{L_1} \cdots R_{L_{n+1}} \sqsubseteq^*_{\mathcal{R}_1} R_L$, which was required to show.

We are now in a position to define the conservative extension $\mathcal{R}'$ of $\mathcal{R}$ required by the theorem. Let $\mathcal{R}'$ be an extension of $\mathcal{R}$ with $\mathcal{R}_1$ and the following axioms for every role $R \in \Sigma$:

$$R \sqsubseteq R_L, \text{ and } R_L \sqsubseteq R, \text{ where } L = L_{\bar{\mathcal{R}}}(R) \in S_{\mathcal{R}}. \tag{26}$$

Clearly, $\mathcal{R}'$ is a conservative extension of R. Indeed, every model $\mathcal{I}$ of $\mathcal{R}$ can be extended to a model of $\mathcal{R}'$ by interpreting the fresh roles R_L for $L \in S_{\mathcal{R}}$ as $R_L^{\mathcal{I}} := \bigcup_{\rho \in L} \rho^{\mathcal{I}}$. It is readily checked that $\mathcal{I}$ satisfies all RIAs (23)–(26).

Before proving that $\bar{\mathcal{R}}'$ is stratified, first note that for every $R_1 \cdots R_n \sqsubseteq R \in \mathcal{R}$ we have $R_{L_1} \cdots R_{L_n} \sqsubseteq^*_{\mathcal{R}_1} R_L$ when $L_i = L_{\bar{\mathcal{R}}}(R_i)$, $(1 \leq i \leq n)$, and $L = L_{\bar{\mathcal{R}}}(R)$. This follows directly from Claim 2 since $L_1 \cdots L_n \subseteq L$. Thus, every RIA in $\mathcal{R}$ is provable using axioms (23)–(26), and so, all RIAs in $\mathcal{R}$ can be disregarded. Moreover, we can regard every role R_L for $L = L_{\bar{\mathcal{R}}}(R)$ as a syntactic variant for the role R because of the axioms (26). Thus, it is sufficient to show that $\bar{\mathcal{R}}_1$ is stratified w.r.t. every $\precsim$ admissible for $\bar{\mathcal{R}}_1$.

Since the RIAs in $\mathcal{R}_1$ do not contain inverses, it is easy to see that $\rho \sqsubseteq^*_{\bar{\mathcal{R}}_1} R$ iff either $\rho \sqsubseteq^*_{\mathcal{R}_1} R$ or $\mathsf{Inv}(\rho) \sqsubseteq^*_{\mathcal{R}_1} \mathsf{Inv}(R)$. So, it is sufficient to show that for every $L_1, \ldots, L_n, L \in S_{\mathcal{R}}$ such that $R_{L_1} \cdots R_{L_n} \sqsubseteq^*_{\mathcal{R}_1} R_L$ and every $k = 1 \ldots n$, there exist R_k^1 and R_k^2 such that:

$$R_{L_1} \cdots R_{L_k} \sqsubseteq^*_{\mathcal{R}_1} R_k^1, \qquad R_k^1 R_{L_{k+1}} \cdots R_{L_n} \sqsubseteq^*_{\mathcal{R}_1} R_L, \tag{27}$$

$$R_{L_k} \cdots R_{L_n} \sqsubseteq^*_{\mathcal{R}_1} R_k^2, \qquad R_{L_1} \cdots R_{L_{k-1}} R_k^2 \sqsubseteq^*_{\mathcal{R}_1} R_L. \tag{28}$$

Indeed, by Claim 2, $L_1 \cdots L_n \subseteq L$. By Claim 1, there exist $L_k^1, L_k^2 \in S_{\mathcal{R}}$ such that $L_1 \cdots L_k \subseteq L_k^1$, $L_k \cdots L_n \subseteq L_k^2$, $L_k^1 L_{k+1} \cdots L_n \subseteq L$, and $L_1 \cdots L_{k-1} L_k^2 \subseteq L$. By Claim 2 we obtain (27) and (28) for $R_k^1 := R_{L_k^1}$ and $R_k^2 := R_{L_k^2}$. □

6 Related Work and Outlook

Complex RIAs are closely related to *inclusion (interaction) axioms* in *grammar modal logics* $\Box_{i_1} \cdots \Box_{i_n} X \to \Box_{j_1} \cdots \Box_{j_n} X$ [6, 5, 8]. Such axioms often cause undecidability, however Baldoni [5] and Demri [8] found a decidable class called the *regular grammar modal logics*. Demri and de Nivelle [9] gave a decision procedure for this class by a translation into the two-variable guarded fragment. The decision procedure assumes that a regular automata are given as an input of the procedure. When applying these results to ontologies and complex RIAs, such a restriction poses a serious practical problem because the users are unlikely to provide such automata. One proposed solution to this problem, is to use a sufficient syntactic condition for regularity, such as $\prec$-regularity [12, 11]. Another sufficient condition [20] requires *associativity* of RIAs: if $R_1 R_2 \sqsubseteq R_1'$ and $R_1' R_3 \sqsubseteq R'$ then there should be R_2' such that $R_2 R_3 \sqsubseteq R_2'$ and $R_1 R_2' \sqsubseteq R'$. A similar condition was required for completeness of the *ordered chaining calculus* for first-order logic with compositional theories [4]. It is easy to see that associativity is a partial case of our sufficient conditions, when $\precsim$ is a total relation on roles. Therefore, our syntactic condition can be seen as a generalization of both associativity and $\prec$-regularity. Note that Theorem 3 is, in fact, proved for total preorders, and therefore it holds for all preorders.

Theorem 1 can possibly be relevant to several results in language theory identifying regular fragments of context-free languages and semi-Thue systems such as *non-self-embedded languages*, *one-letter grammars*, and *finite languages* (see, e.g., [7, 1, 15]). However, neither the original regularity condition for $\mathcal{SROIQ}$ [11], nor our extended condition, nor the associativity condition seem to relate to the known cases of regular context-free grammars. The reason could probably be that the conditions that are natural for compositional properties of binary relations ($R_1R_2 \sqsubseteq R_3$) might be not be so natural in the context of formal language theory ($A_1 \to A_2A_3$) and vice versa.

Theorem 3 means that stratified sets of RIAs can express any regular compositional properties of roles. In other words, our syntactic restriction has already maximal expressive power w.r.t. such properties and no further extension is necessary. Note that the proof of Theorem 3 is not constructive: it does not provide an algorithm for building the extension $\mathcal{R}'$ *automatically* from $\mathcal{R}$—it is necessary to know the regular automata for $L_{\bar{\mathcal{R}}}(R)$. It is an interesting question whether there exists such a completion procedure that terminates if all $L_{\bar{\mathcal{R}}}(R)$ are regular. It seems to be not even clear if it is possible to effectively check regularity for $\mathcal{R}$. It was claimed [9] that this problem is undecidable since it is undecidable whether a linear grammar is regular. But the problem of regularity for context-free grammars seems to be harder since context-free grammars distinguish between terminal and non-terminal symbols. There is no such a distinction between types of roles in RIAs, which makes a reduction from context-free grammars to sets of RIAs problematic. In this respect, the sets of RIAs are more related to so-called *sentential forms of context-free grammars* [16] or *pure context-free grammars* [14] where the symbols are not distinguished. A sentential form of a context-free grammar is a pure grammar that generates the language consisting of terminal and non-terminal symbols. The resulted language can be non-regular even for a regular grammar. For example, the linear grammar $A \to a$, $A \to aAa$ generates a regular language $L(A) = a(aa)^*$, but its sentential form generates a non-regular language $L^s(A) = L(A) \cup \{a^iAa^i \mid i \geq 0\}$. Pure grammars have different algorithmic properties than context-free grammars. For example, unlike for context-free grammars, given a pure context-free grammar and a regular automaton, it is decidable if they generate the same language [14]. The problem of regularity for pure grammars, however, is still open, to the best of our knowledge.

In this work we introduced a notion of stratified set of role inclusion axioms which provides a syntactically-checkable sufficient condition for regularity of RIAs—a condition that ensures decidability of $\mathcal{SROIQ}$ [11]. We demonstrated that for every stratified $\mathcal{SROIQ}$ ontology, one can construct a regular automaton representing the RIAs, which is worst case exponential in the size of the ontology. This implies that the complexity of reasoning with extended $\mathcal{SROIQ}$ remains the same as the complexity of the original $\mathcal{SROIQ}$, namely N2ExpTime-complete [13]. Moreover, we demonstrated that our conditions for regularity are in a sense maximal—every ontology $\mathcal{O}$ with regular RIAs can be conservatively extended to an ontology with stratified RIAs.

Acknowledgement

The author would like to thank the anonymous referees for valuable comments and Frantisek Simancik for pointing out to the literature on sentential forms and pure grammars. This work was supported by the EPSRC, grant number EP/G02085X.

References

1. Andrei, S., Chin, W.N., Cavadini, S.V.: Self-embedded context-free grammars with regular counterparts. Acta Inf. 40(5), 349–365 (2004)
2. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proc. IJCAI-05. pp. 364–369 (2005)
3. Baader, F., Nipkow, T.: Term rewriting and all that. Cambridge University Press, New York, NY, USA (1998)
4. Bachmair, L., Ganzinger, H.: Ordered chaining calculi for first-order theories of transitive relations. J. ACM 45(6), 1007–1049 (1998)
5. Baldoni, M.: Normal Multimodal Logics: Automatic Deduction and Logic Programming Extension. Ph.D. thesis, Università degli Studi di Torino (1998)
6. Fariñas del Cerro, L., Penttonen, M.: Grammar logics. Logique et Analyse 121-122, 123–134 (1988)
7. Chomsky, N.: On certain formal properties of grammars. Information and Control 2(2), 137–167 (1959)
8. Demri, S.: The complexity of regularity in grammar logics and related modal logics. J. Log. Comput. 11(6), 933–960 (2001)
9. Demri, S., de Nivelle, H.: Deciding regular grammar logics with converse through first-order logic. Journal of Logic, Language and Information 14(3), 289–329 (2005)
10. Hopcroft, J.E., Ullman, J.D.: Introduction to Automata Theory, Languages and Computation. Addison-Wesley (1979)
11. Horrocks, I., Kutz, O., Sattler, U.: The even more irresistible $\mathcal{SROIQ}$. In: KR. pp. 57–67 (2006)
12. Horrocks, I., Sattler, U.: Decidability of $\mathcal{SHIQ}$ with complex role inclusion axioms. Artif. Intell. 160(1-2), 79–104 (2004)
13. Kazakov, Y.: $\mathcal{RIQ}$ and $\mathcal{SROIQ}$ are harder than $\mathcal{SHOIQ}$. In: KR. pp. 274–284. AAAI Press (2008)
14. Maurer, H., Salomaa, A., Wood, D.: Pure grammars. Information and Control 44(1), 47 – 72 (1980)
15. Nederhof, M.J.: Practical experiments with regular approximation of context-free languages. Computational Linguistics 26(1), 17–44 (2000)
16. Salomaa, A.: On sentential forms of context-free grammars. Acta Informatica 2(1), 40–49 (1973)
17. Schulz, S., Romacker, M., Hahn, U.: Part-whole reasoning in medical ontologies revisited: Introducing SEP triplets into classification-based description logics. In: Proc. of the 1998 AMIA Annual Fall Symposium. pp. 830–834. Hanley & Belfus (1998)
18. Sipser, M.: Introduction to the Theory of Computation, Second Edition. Course Technology (February 2005)
19. Suntisrivaraporn, B., Baader, F., Schulz, S., Spackman, K.A.: Replacing SEP-triplets in SNOMED CT using tractable description logic operators. In: AIME. Lecture Notes in Computer Science, vol. 4594, pp. 287–291. Springer (2007)
20. Wessel, M.: Obstacles on the way to qualitative spatial reasoning with description logics: Some undecidability results. In: Description Logics. CEUR Workshop Proceedings, vol. 49. CEUR-WS.org (2001)