

# Reachability modules for the Description Logic *SRIQ*

Riku Nortje, Katarina Britz, and Thomas Meyer

Center for Artificial Intelligence Research, University of KwaZulu-Natal and CSIR  
Meraka Institute, South Africa  
Email: {rnortje;abritz;tmeyer}@csir.co.za

**Abstract.** In this paper we investigate module extraction for the Description Logic *SRIQ*. We formulate modules in terms of the reachability problem for directed hypergraphs. Using inseparability relations, we investigate the module-theoretic properties of reachability modules and show by means of an empirical evaluation that these modules have the potential of being substantially smaller than syntactic locality modules.

## 1 Introduction

Description Logics (DLs) are widely used in ontological modeling. They form a family of knowledge representation languages that are mostly decidable fragments of first-order logic. Their formal semantics not only allow for the exchange of DL ontologies but provide support for reasoning — the computation of additional logical inferences from the facts stated explicitly in an ontology.

There are many different DLs, each differing in the expressivity of the language and the complexity of reasoning. In general, the more expressive a DL the more complex the reasoning associated with it. This allows the ontology modeller to choose, for the intended application, the best balance between language expressivity on the one hand and reasoning complexity on the other. The DL *SRIQ* is an expressive language and is a subset *SROIQ*, the W3C OWL DL Web Ontology language.

Modularization plays an important part in the design and maintenance of large scale ontologies. Modules are loosely defined as subsets of ontologies that cover some topic of interest, where the topic of interest is defined by a set of symbols. Extracting minimal modules is computationally expensive and even undecidable for expressive DLs [4, 5]. Therefore, the use of approximation techniques and heuristics play an important role in the efficient design of algorithms.

Syntactic locality [4, 5], because of its excellent model-theoretic properties, has become an ideal heuristic and is widely used in a diverse set of algorithms [19, 3, 6]. Suntisrivaraporn [19] showed that, for the DL  $\mathcal{EL}^+$ ,  $\perp$ -locality module extraction is equivalent to the reachability problem in directed hypergraphs. Nortjé et al. [14, 15] extended the reachability problem to include  $\top$ -locality and introduced bidirectional reachability modules as a subset of  $\perp\top^*$ -locality modules. This work was further extended to the DL *SROIQ* by Nortje et al.

[16] who showed that extracting  $\perp\top^*$ -reachability modules is equivalent to extracting frontier graphs in hypergraphs. Reachability modules are not only of importance in hypergraph-based reasoning support for CBoxes [16], but are potentially smaller than syntactic locality modules.

In this paper we investigate the module-theoretic properties of reachability modules for the DL  $\mathcal{SRIQ}$ . We show that these modules are not self-contained or depleting but they are robust under vocabulary restrictions, vocabulary extensions, replacement and joins. By showing that reachability modules preserve all justifications for entailments, we show that depleting modules are sufficient for preserving all justifications but not necessary. This paper is an extended version of the paper presented at DL2013 [17].

In Section 2 we give a brief introduction to the DL  $\mathcal{SRIQ}$ , hypergraphs and modularization as defined by inseparability relations. Section 3 introduces a normal form for  $\mathcal{SRIQ}$  CBoxes as well as the rules necessary to transform any such CBox to normal form. In Section 4 we introduce both  $\perp$ - and  $\top$  reachability modules and investigate all their module theoretic properties in terms of inseparability relations. All proofs of the work presented appear in the accompanying appendix. In Section 5 we show the results of an empirical evaluation of these modules. Lastly in Section 6 we conclude this paper with a short summary of the results.

## 2 Background

In Section 2.1 we give a brief introduction to DLs and modularization with specific focus on the DL  $\mathcal{SRIQ}$  [9]. In Section 2.2 we give a brief introduction to modules and module theoretic properties.

### 2.1 The DL $\mathcal{SRIQ}$

The syntax and semantics of  $\mathcal{SRIQ}$  is listed in Table 2.1.  $N_C$  and  $N_R$  denote disjoint sets of atomic concept names and role names. The set  $N_R$  includes the universal role whilst  $N_C$  excludes the  $\top$  and  $\perp$  concepts. For a complete definition of  $\mathcal{SRIQ}$ , refer to Horrocks et al. [9], and for Description Logics refer to [2].

In order to ensure decidability in  $\mathcal{SRIQ}$  there are some restrictions on the use of roles.  $R_1 \circ \dots \circ R_n \sqsubseteq R$ , where  $n \geq 1$  and  $R_i, R \in N_R$ , is a *role inclusion axiom* (RIA). A *role hierarchy* is a finite set of RIAs. Here  $R_1 \circ \dots \circ R_n$  denotes a composition of roles where  $R, R_i$  may also be an *inverse role*  $R^-$ . A role  $R$  is *simple* if (i) it does not appear on the right-hand side of a RIA, or (ii) is the inverse of a simple role, or (iii) appears on the right-hand side of a RIA only if the left-hand side is a simple role.  $Ref(R)$ ,  $Irr(R)$  and  $Dis(R, S)$ , where  $R, S$  are roles other than  $U$ , are role assertions. A set of role assertions is simple w.r.t. a role-hierarchy  $H$  if each assertion  $Irr(R)$  and  $Dis(R, S)$  uses only simple roles w.r.t.  $H$ .

A strict partial order  $\prec$  on  $N_R$  is a *regular order* if, and only if, for all roles  $R$  and  $S$ :  $S \prec R$  iff  $S^- \prec R$ . Let  $\prec$  be a regular order on roles. A RIA  $w \sqsubseteq R$  is

Constructs	Syntax	Semantics
atomic concept	$C$	$C^{\mathcal{I}} \in \Delta^{\mathcal{I}}, C \in N_C$
role	$R$	$R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}, R \in N_R$
inverse role	$R^{-}$	$R^{-\mathcal{I}} = \{(y, x) \mid (x, y) \in R^{\mathcal{I}}\}, R \in N_R$
universal role	$U$	$U^{\mathcal{I}} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
role composition	$R_1 \circ \dots \circ R_n$	$\{(x, z) \mid (x, y_1) \in R_1^{\mathcal{I}} \wedge (y_1, y_2) \in R_2^{\mathcal{I}} \wedge \dots \wedge (y_n, z) \in R_n^{\mathcal{I}}, n \geq 2, R_i \in N_R\}$
top	$\top$	$\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$
bottom	$\perp$	$\perp^{\mathcal{I}} = \emptyset$
negation	$\neg C$	$(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
conjunction	$C_1 \sqcap C_2$	$(C_1 \sqcap C_2)^{\mathcal{I}} = C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$
disjunction	$C_1 \sqcup C_2$	$(C_1 \sqcup C_2)^{\mathcal{I}} = C_1^{\mathcal{I}} \cup C_2^{\mathcal{I}}$
exist restriction	$\exists R.C$	$\{x \mid (\exists y)[(x, y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}]\}$
value restriction	$\forall R.C$	$\{x \mid (\forall y)[(x, y) \in R^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}}]\}$
self restriction	$\exists R.Self$	$\{x \mid (x, x) \in R^{\mathcal{I}}\}$
atmost restriction	$\leq nR.C$	$\{x \mid \#\{y \mid (x, y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\} \leq n\}$
atleast restriction	$\geq nR.C$	$\{x \mid \#\{y \mid (x, y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\} \geq n\}$
Axiom	Syntax	Semantics
concept inclusion	$C_1 \sqsubseteq C_2$	$C_1^{\mathcal{I}} \subseteq C_2^{\mathcal{I}}$
role inclusion	$R_1 \circ \dots \circ R_n \sqsubseteq R_{n+1}$	$(R_1 \circ \dots \circ R_n)^{\mathcal{I}} \subseteq R_{n+1}^{\mathcal{I}}, n \geq 1$
reflexivity	$Ref(R)$	$\{(x, x) \mid x \in \Delta^{\mathcal{I}}\} \subseteq R^{\mathcal{I}}$
irreflexivity	$Irr(R)$	$\{(x, x) \mid x \in \Delta^{\mathcal{I}}\} \cap R^{\mathcal{I}} = \emptyset$
disjointness	$Dis(R, S)$	$S^{\mathcal{I}} \cap R^{\mathcal{I}} = \emptyset$

**Table 1.** Syntax and semantics of  $\mathcal{SRIQ}$

$\prec$ -regular if, and only if,  $R \in N_R$  and  $w$  has one of the following forms:  $R \circ R$ ;  $R^{-}$ ;  $S_1 \circ \dots \circ S_n$ , where each  $S_i \prec R$ ;  $R \circ S_1 \circ \dots \circ S_n$ , where each  $S_i \prec R$  or  $S_1 \circ \dots \circ S_n \circ R$ , where each  $S_i \prec R$ . A role hierarchy  $H$  is *regular* if there exists a regular order  $\prec$  such that each RIA in  $H$  is  $\prec$ -regular. An  $RBox$  is a finite, regular role hierarchy  $H$  together with a finite set of role assertions simple w.r.t.  $H$ .

The set of  $\mathcal{SRIQ}$  *concept descriptions* is the smallest set such that:

1.  $\perp, \top$ , and each  $C \in N_C$  is a concept description.
2. If  $C$  is a concept description, then  $\neg C$  is a concept description.
3. If  $C$  and  $D$  are concept descriptions,  $R$  is a role,  $S$  is a simple role, and  $n$  is a non-negative integer, then the following are all concept descriptions:

$$(C \sqcap D), (C \sqcup D), \exists R.C, \forall R.C, \leq nS.C, \geq nS.C, \exists S.Self$$

If  $C$  and  $D$  are concept description then  $C \sqsubseteq D$  is a *general concept inclusion* (GCI) axiom. A  $TBox$  is a finite set of GCIs. If  $C$  is a concept description,  $a, B \in N_I$ ,  $R, S \in N_R$  with  $S$  a simple role, then  $C(a)$ ,  $R(a, b)$ ,  $\neg S(a, b)$ , and  $a \neq b$ , are individual assertions. An  $\mathcal{SRIQ}$   $ABox$  is a finite set of individual assertions. All GCIs, RIAs, role assertions, and individual assertions are referred to as axioms. A  $\mathcal{SRIQ}$ -KB base is the union of a  $TBox$ ,  $RBox$  and  $ABox$ . Given a  $\mathcal{SRIQ}$   $TBox$   $\mathcal{T}$  and  $RBox$   $\mathcal{R}$  we define a  $\mathcal{SRIQ}$   $CBox$   $\mathcal{C}$  as  $\mathcal{T} \cup \mathcal{R}$ .

## 2.2 Modules and their properties

Module extraction is the process of extracting subsets of axioms from CBoxes that are self contained with respect to some criteria. These sets of axioms, called *modules*, may be used for various purposes such as reuse, optimization and error pinpointing amongst others [5, 19].

**Definition 1. (Module for the arbitrary DL  $\mathcal{L}$  [11, 12])** *Let  $\mathcal{L}$  be an arbitrary description language,  $\mathcal{O}$  an  $\mathcal{L}$  ontology, and  $\sigma$  a statement formulated in  $\mathcal{L}$ . Then,  $\mathcal{O}' \subseteq \mathcal{O}$  is a module for  $\sigma$  in  $\mathcal{O}$  (a  $\sigma$ -module in  $\mathcal{O}$ ) whenever:  $\mathcal{O} \models \sigma$  if and only if  $\mathcal{O}' \models \sigma$ .*

Definition 1 is sufficiently general so that any subset of an ontology preserving a statement of interest is considered a module, the entire ontology is therefore a module in itself.

Different use cases usually result in different notions of what the definition and characteristics of a module should be. Modules are often defined via the notion of conservative extensions. Given some signature (a set of concept and role names) and a set of axioms, a conservative extension of this set is simply one that implies all the same consequences over the signature. More formally:

**Definition 2. (Conservative extension [1, 7])** *Let  $\mathcal{C}$  and  $\mathcal{C}_1$  be two CBoxes such that  $\mathcal{C}_1 \subseteq \mathcal{C}$ , and let  $\Sigma$  be a signature. Then*

- $\mathcal{C}$  is a  $\Sigma$ -conservative extension of  $\mathcal{C}_1$  if, for every  $\alpha$  with  $\text{Sig}(\alpha) \subseteq \Sigma$ , we have  $\mathcal{C} \models \alpha$  iff  $\mathcal{C}_1 \models \alpha$ .
- $\mathcal{C}$  is a conservative extension of  $\mathcal{C}_1$  if  $\mathcal{C}$  is a  $\Sigma$ -conservative extension of  $\mathcal{C}_1$  for  $\Sigma = \text{Sig}(\mathcal{C}_1)$ .

Given that both sets of axioms imply the same consequences for a given signature we may then use the smaller set whenever we wish to reason over this signature. A closely related notion to conservative extensions is that of *inseparability*.

**Definition 3.** [18]  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are  $\Sigma$ -concept name inseparable, written  $\mathcal{C}_1 \equiv_{\Sigma}^c \mathcal{C}_2$ , if for all  $\Sigma$ -concept names  $C, D$ , it holds that  $\mathcal{C}_1 \models C \sqsubseteq D$  if and only if  $\mathcal{C}_2 \models C \sqsubseteq D$ .

**Definition 4.** [18]  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are  $\Sigma$ -subsumption inseparable, written  $\mathcal{C}_1 \equiv_{\Sigma}^s \mathcal{C}_2$ , if for all terms  $X, Y$  that are concepts or roles over  $\Sigma$ , it holds that  $\mathcal{C}_1 \models X \sqsubseteq Y$  if and only if  $\mathcal{C}_2 \models X \sqsubseteq Y$ .

**Definition 5.** [11, 12, 18] Let  $\mathcal{C}$  be a CBox,  $\mathcal{M} \subseteq \mathcal{C}$ ,  $S$  an inseparability relation and  $\Sigma$  a signature. We call  $\mathcal{M}$

- an  $S_{\Sigma}$ -module of  $T$  if  $\mathcal{M} \equiv_{\Sigma}^S \mathcal{C}$ .
- a self-contained  $S_{\Sigma}$ -module of  $\mathcal{C}$  if  $\mathcal{M} \equiv_{\Sigma \cup \text{Sig}(\mathcal{M})}^S \mathcal{C}$ .
- a depleting  $S_{\Sigma}$ -module of  $\mathcal{C}$  if  $\emptyset \equiv_{\Sigma \cup \text{Sig}(\mathcal{M})}^S \mathcal{C} \setminus \mathcal{M}$ .

Modules may therefore be characterized by some inseparability criteria. It is of interest how modules defined this way would behave under different use case scenarios. For this purpose, several properties of inseparability relations [10] have been investigated in the literature, which allows us to compare different definitions of modules. Given a CBox  $\mathcal{C}$  and a module  $\mathcal{M} \subseteq \mathcal{C}$  for a signature  $\Sigma$ , we are interested in the following inseparability properties:

- *Robustness under vocabulary restrictions* implies that when we wish to restrict the symbols from  $\Sigma$  further we do not need to import a different module and may continue to use  $\mathcal{M}$ .
- *Robustness under vocabulary extension* implies that should we wish to add new symbols to  $\Sigma$  that do not appear in  $\mathcal{C}$  we do not need to use a different module but may use  $\mathcal{M}$ .
- *Robustness under replacement* ensures that the result of importing  $\mathcal{M}$  into a CBox  $\mathcal{C}_1$  is a module of the result of importing  $\mathcal{C}$  into  $\mathcal{C}_1$ . This is also called module coverage and refers to the fact that importing a module does not affect its property of being a module.
- *Robustness under joins* implies that if  $\mathcal{C}$  and  $\mathcal{C}_1$  are inseparable w.r.t.  $\Sigma$  and all the terms they share are from  $\Sigma$ , then each of them are inseparable with their union w.r.t.  $\Sigma$ .

More formally:

**Definition 6.** [10–12] *The inseparability relation  $S$  is called*

- *robust under vocabulary restrictions if, for all CBoxes  $\mathcal{C}_1, \mathcal{C}_2$  and all signatures  $\Sigma, \Sigma'$  with  $\Sigma \subseteq \Sigma'$ , the following holds: if  $\mathcal{C}_1 \equiv_{\Sigma}^S \mathcal{C}_2$ , then  $\mathcal{C}_1 \equiv_{\Sigma'}^S \mathcal{C}_2$ .*
- *robust under vocabulary extensions if, for all CBoxes  $\mathcal{C}_1, \mathcal{C}_2$  and all signatures  $\Sigma, \Sigma'$  with  $\Sigma' \cap (\text{Sig}(\mathcal{C}_1) \cup \text{Sig}(\mathcal{C}_2)) \subseteq \Sigma$ , the following holds: if  $\mathcal{C}_1 \equiv_{\Sigma}^S \mathcal{C}_2$ , then  $\mathcal{C}_1 \equiv_{\Sigma'}^S \mathcal{C}_2$ .*
- *robust under replacement if, for all CBoxes  $\mathcal{C}_1, \mathcal{C}_2$  and all signatures  $\Sigma$  and every CBox  $\mathcal{C}$  with  $\text{Sig}(\mathcal{C}) \cap (\text{Sig}(\mathcal{C}_1) \cup \text{Sig}(\mathcal{C}_2)) \subseteq \Sigma$ , the following holds: if  $\mathcal{C}_1 \equiv_{\Sigma}^S \mathcal{C}_2$  then  $\mathcal{C}_1 \cup \mathcal{C} \equiv_{\Sigma}^S \mathcal{C}_2 \cup \mathcal{C}$ .*
- *robust under joins if, for all CBoxes  $\mathcal{C}_1, \mathcal{C}_2$  and all signatures  $\Sigma$  with  $\text{Sig}(\mathcal{C}) \cap \text{Sig}(\mathcal{C}_2) \subseteq \Sigma$ , if  $\mathcal{C}_1 \equiv_{\Sigma}^S \mathcal{C}_2$  then  $\mathcal{C}_i \equiv_{\Sigma}^S \mathcal{C}_1 \cup \mathcal{C}_2$ , for  $i = 1, 2$ .*

Deciding conservative extensions has been shown to be computationally expensive or even undecidable for relatively inexpressive DLs. Therefore, an approximation of these modules, based on syntax, called syntactic locality modules has been introduced [5]. Syntactic locality modules possess all the module-theoretic properties discussed in this section and have become one of the most widely used definitions of modules. We will give a definition of a normalized version of syntactic locality once we have introduced a normal form for *SRIQ*.

### 3 Normal Form

In this section we introduce a normal form for *SRIQ* CBoxes. We utilize normalization in order to simplify the definitions, to ease the understanding of the work that follows, as well as to simplify the presentation of proofs.

**Definition 7.** Given  $B_i \in (N_C \cup \{\top\})$ ,  $C_i \in (N_C \cup \{\perp\})$ ,  $D \in \{\exists R.B, \geq nR.B, \exists R.Self\}$ , with  $R, S, R_i, S_i$  role names from  $N_R$  or their inverses and  $n \geq 1$ , a *SRIQ* CBox  $\mathcal{C}$  is in **normal form** if every axiom  $\alpha \in \mathcal{C}$  is in one of the following forms:

$\alpha_1: B_1 \sqcap \dots \sqcap B_n \sqsubseteq C_1 \sqcup \dots \sqcup C_m$	$\alpha_2: D \sqsubseteq C_1 \sqcup \dots \sqcup C_m$
$\alpha_3: B_1 \sqcap \dots \sqcap B_n \sqsubseteq D$	$\alpha_4: R_1 \circ \dots \circ R_n \sqsubseteq R_{n+1}$
$\alpha_5: R_1 \sqsubseteq R_2$	$\alpha_6: D_1 \sqsubseteq D_2$
$\alpha_7: Dis(R_1, R_2)$	

In order to normalize a *SRIQ* CBox  $\mathcal{C}$  we repeatedly apply the normalization rules from Table 2. Each application of a rule rewrites an axiom into its equivalent normal form. It is easy to see that the application of every rule ensures that the normalized CBox is a conservative extension of the original. We note that the *SRIQ* axiom  $Ref(R)$  is represented by its equivalent  $\top \sqsubseteq \exists R.Self$  and  $Irr(R)$  by  $\exists R.Self \sqsubseteq \perp$  [2].

**Table 2.** *SRIQ* normalization rules

NR1	$\hat{B} \sqcap \neg \hat{C}_2 \sqsubseteq \hat{C}_1 \rightsquigarrow \hat{B} \sqsubseteq \hat{C}_1 \sqcup \hat{C}_2$
NR2	$\hat{B}_1 \sqsubseteq \hat{C} \sqcup \neg \hat{B}_2 \rightsquigarrow \hat{B}_1 \sqcap \hat{B}_2 \sqsubseteq \hat{C}$
NR3	$\hat{B} \sqcap \hat{D} \sqsubseteq \hat{C} \rightsquigarrow \hat{B} \sqcap A \sqsubseteq \hat{C}, \hat{D} \sqsubseteq A, A \sqsubseteq \hat{D}$
NR4	$\hat{B} \sqsubseteq \hat{C} \sqcup \hat{D} \rightsquigarrow \hat{B} \sqsubseteq \hat{C} \sqcup A, \hat{D} \sqsubseteq A, A \sqsubseteq \hat{D}$
NR5	$\hat{B} \sqsubseteq \hat{C}_1 \sqcap \hat{C}_2 \rightsquigarrow \hat{B} \sqsubseteq \hat{C}_1, \hat{B} \sqsubseteq \hat{C}_2$
NR6	$\hat{B}_1 \sqcup \hat{B}_2 \sqsubseteq \hat{C} \rightsquigarrow \hat{B}_1 \sqsubseteq \hat{C}, \hat{B}_2 \sqsubseteq \hat{C}$
NR7	$\dots \forall R. \hat{C} \dots \rightsquigarrow \dots \neg \exists R. A \dots, A \sqcap \hat{C} \sqsubseteq \perp, \top \sqsubseteq A \sqcup \hat{C}$
NR8	$\dots \exists R. \hat{D} \dots \rightsquigarrow \dots \exists R. A \dots, \hat{D} \sqsubseteq A, A \sqsubseteq \hat{D}$
NR9	$\dots \geq nR. \hat{D} \dots \rightsquigarrow \dots \geq nR. A \dots, \hat{D} \sqsubseteq A, A \sqsubseteq \hat{D}$
NR10	$\dots \leq nR. \hat{C} \dots \rightsquigarrow \dots \neg (\geq (n+1)R. \hat{C}) \dots$
NR11	$\hat{B} \equiv \hat{C} \rightsquigarrow \hat{B} \sqsubseteq \hat{C}, \hat{C} \sqsubseteq \hat{B}$
NR12	$\geq 0R. B \sqsubseteq \hat{C} \rightsquigarrow \top \sqsubseteq \hat{C}$
NR13	$\hat{B} \sqsubseteq \exists R. \perp \rightsquigarrow \hat{B} \sqsubseteq \perp$
NR14	$\hat{B} \sqsubseteq \geq nR. \perp \rightsquigarrow \hat{B} \sqsubseteq \perp$
NR15	$\hat{B} \sqsubseteq \geq 0R. B \rightsquigarrow$
NR16	$\geq nR. \perp \sqsubseteq \hat{C} \rightsquigarrow$
NR17	$\exists R. \perp \sqsubseteq \hat{C} \rightsquigarrow$
NR18	$\hat{B} \sqcap \perp \sqsubseteq \hat{C} \rightsquigarrow$
NR19	$\perp \sqsubseteq \hat{C} \rightsquigarrow$
NR20	$\hat{B} \sqsubseteq \hat{C} \sqcup \top \rightsquigarrow$
NR21	$\hat{B} \sqsubseteq \top \rightsquigarrow$

Above  $A$  is a new concept name not in  $N_C$ ,  $\hat{B}_i$  and  $\hat{C}_i$  are possibly complex concept descriptions and  $\hat{D}$  a complex concept description.  $R \in N_R$  or it's inverse,  $n \geq 0$

**Theorem 1.** Exhaustively applying the rules from Table 2 to any *SRIQ* CBox  $\mathcal{C}$  results in a *SRIQ* CBox  $\mathcal{C}'$  in normal form. The normalization process can be completed in linear time in the number of axioms.

**Proof Sketch:** We show that normalization is linear in the number of axioms by applying normalization rules in the following order:  $\equiv$ -elimination (NR11),  $\forall$ -elimination (NR7),  $\leq$ -elimination (NR10), Complex role-filler elimination (NR8, NR9),  $\neg$ -elimination and simplification by iteration of rules NR1, NR3, NR6 and NR2, NR4, NR5. Lastly rules NR12 through NR21 are applied.  $\square$

*Example 1.* Let  $\alpha_1 = B \sqsubseteq \neg C$ , and  $\alpha_2 = \neg A \sqsubseteq B$ . Then,  $\alpha_1$  may be normalized by application of rule NR2 to  $\alpha_1^N = B \sqcap C \sqsubseteq \perp$  since  $\neg C = \neg C \sqcup \perp$ .  $\alpha_2$  may be normalized by application of rule NR1 to  $\alpha_2^N = \top \sqsubseteq B \cup A$  since  $\neg A = \neg A \sqcap \top$ .

We will discuss the importance of normalization in the context of this paper in more detail in the next section.

## 4 Reachability Modules

Syntactic locality is a widely used approximation to deciding conservative extensions. Given a normalized CBox  $\mathcal{C}$ , the definition of syntactic locality can be simplified to the following:

**Definition 8. (Normalized Syntactic Locality)** *Let  $\Sigma$  be a signature and  $\mathcal{C}$  a normalized  $\mathcal{SRIQ}$  CBox. An axiom  $\alpha$  is  $\perp$ -local w.r.t.  $\Sigma$  ( $\top$ -local w.r.t.  $\Sigma$ ) if  $\alpha \in \text{Ax}(\Sigma)^\perp$  ( $\alpha \in \text{Ax}(\Sigma)^\top$ ), as defined in the grammar:*

$$\begin{array}{l}
 \hline
 \perp\text{-Locality} \\
 \hline
 \text{Ax}(\Sigma)^\perp ::= C^\perp \sqsubseteq C \mid w^\perp \sqsubseteq R \mid \text{Dis}(S^\perp, S) \mid \text{Dis}(S, S^\perp) \\
 \text{Con}^\perp(\Sigma) ::= A^\perp \mid C^\perp \sqcap C \mid C \sqcap C^\perp \mid \exists R^\perp.C \mid \exists R.C^\perp \mid \exists R^\perp.\text{Self} \mid \\
 \qquad \qquad \qquad \geq nR^\perp.C \mid \geq nR.C^\perp \\
 \hline
 \top\text{-Locality} \\
 \hline
 \text{Ax}(\Sigma)^\top ::= C \sqsubseteq C^\top \mid w \sqsubseteq R^\top \\
 \text{Con}^\top(\Sigma) ::= A^\top \mid C^\top \sqcup C \mid C \sqcup C^\top \mid \exists R^\top.C^\top \mid \geq nR^\top.C^\top \mid \\
 \qquad \qquad \qquad \exists R^\top.\text{Self} \\
 \hline
 \end{array}$$

In the grammar, we have that  $A^\perp, A^\top \notin \Sigma$  is an atomic concept,  $R^\perp$  (resp.  $S^\perp$ ) is either an atomic role (resp. a simple atomic role) not in  $\Sigma$  or the inverse of an atomic role (resp. of a simple atomic role) not in  $\Sigma$ ,  $C$  is any concept,  $R$  is any role,  $S$  is any simple role, and  $C^\perp \in \text{Con}^\perp(\Sigma)$ ,  $C^\top \in \text{Con}^\top(\Sigma)$ . We also denote by  $w^\perp$  a role chain  $w = R_1 \circ \dots \circ R_n$  such that for some  $i$  with  $1 \leq i \leq n$ , we have that  $R_i$  is (possibly inverse of) an atomic role not in  $\Sigma$ . A CBox  $\mathcal{C}$  is  $\perp$ -local ( $\top$ -local) w.r.t.  $\Sigma$  if  $\alpha$  is  $\perp$ -local ( $\top$ -local) w.r.t.  $\Sigma$  for all  $\alpha \in \mathcal{C}$ .

For a complete overview of locality modules as well as algorithms for extracting such we refer the interested reader to Cuenca Grau et al [5].

A variant of  $\perp$ -syntactic locality modules called  $\perp$ -reachability based modules [19] is based on the reachability problem in directed hypergraphs. Hypergraphs [13, 20] are a generalization of graphs and have been studied extensively since the 1970s as a powerful tool for modeling many problems in Discrete Mathematics.

We extend the work done by Nortje et al.[15] and define reachability for *SRIQ* CBoxes. We then continue to show that these modules share all the robustness properties of locality modules and therefore is well suited to be used in the ontology reuse scenario.

**Definition 9. ( $\perp$ -Reachability)** *Let  $\mathcal{C}$  be a *SRIQ* CBox in normal form and  $\Sigma \subseteq \text{Sig}(\mathcal{C})$  a signature. The set of  $\perp$ -reachable names in  $\mathcal{C}$  w.r.t.  $\Sigma$ , denoted by  $\Sigma_{\mathcal{C}}^{\leftarrow \perp}$ , is defined inductively as follows:*

- For every  $x \in (\Sigma \cup \{\top\})$  we have  $x \in \Sigma_{\mathcal{C}}^{\leftarrow \perp}$ .
- For every inclusion axiom  $(\alpha_L \sqsubseteq \alpha_R) \in \mathcal{C}$ , if  $\text{Sig}(\alpha_L) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow \perp}$  then every  $y \in \text{Sig}(\alpha_R)$  is also in  $\Sigma_{\mathcal{C}}^{\leftarrow \perp}$ .

*Every axiom  $\alpha := \alpha_L \sqsubseteq \alpha_R$  such that  $\text{Sig}(\alpha_L) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow \perp}$  we call  $\Sigma_{\mathcal{C}}^{\leftarrow \perp}$ -reachable. Axioms of the form  $\text{Dis}(R, S) \in \mathcal{C}$  are  $\Sigma_{\mathcal{C}}^{\leftarrow \perp}$ -reachable whenever  $\{R, S\} \subseteq \Sigma_{\mathcal{C}}^{\leftarrow \perp}$ . The set of all  $\Sigma_{\mathcal{C}}^{\leftarrow \perp}$ -reachable axioms is denoted by  $\mathcal{C}_{\Sigma}^{\leftarrow \perp}$  and is called the  $\perp$ -reachability module for  $\mathcal{C}$  over  $\Sigma$ .*

It is self-evident from Definition 8 that an axiom is  $\perp$ -reachable w.r.t  $\Sigma$  exactly when it is not  $\perp$ -local w.r.t.  $\Sigma$ . Similarly we define an axiom to be  $\top$ -reachable exactly when it is not  $\top$ -local.

**Definition 10. ( $\top$ -Reachability)** *Let  $\mathcal{C}$  be a *SRIQ* CBox in normal form and  $\Sigma \subseteq \text{Sig}(\mathcal{C})$  a signature. The set of  $\top$ -reachable names in  $\mathcal{C}$  w.r.t.  $\Sigma$ , denoted by  $\Sigma_{\mathcal{C}}^{\leftarrow \top}$ , is defined inductively as follows:*

- For every  $x \in (\Sigma \cup \{\perp\})$  we have that  $x \in \Sigma_{\mathcal{C}}^{\leftarrow \top}$ .
- For all inclusion axioms  $(\alpha_L \sqsubseteq \alpha_R) \in \mathcal{C}$ , if
  - $\alpha_R = \perp$ , or
  - $\alpha_R$  is of the form  $A_1 \sqcup \dots \sqcup A_n$  and all  $A_i \in \Sigma_{\mathcal{C}}^{\leftarrow \top}$ , or
  - $\alpha_R$  has any other form and there exists some  $x \in \text{Sig}(\alpha_R) \cap \Sigma_{\mathcal{C}}^{\leftarrow \top}$
 then every  $y \in \text{Sig}(\alpha_L)$  is also in  $\Sigma_{\mathcal{C}}^{\leftarrow \top}$ .

*Every axiom  $\alpha := \alpha_L \sqsubseteq \alpha_R$  such that,  $\alpha_R = \perp$ , or  $\alpha_R$  is of the form  $A_1 \sqcup \dots \sqcup A_n$  and all  $A_i \in \Sigma_{\mathcal{C}}^{\leftarrow \top}$ , or  $\alpha_R$  has any other form and there exists some  $x \in \text{Sig}(\alpha_R) \cap \Sigma_{\mathcal{C}}^{\leftarrow \top}$ , we call  $\Sigma_{\mathcal{C}}^{\leftarrow \top}$ -reachable. All axioms of the form  $\text{Dis}(R, S) \in \mathcal{C}$  are always  $\Sigma_{\mathcal{C}}^{\leftarrow \top}$ -reachable and  $\{R, S\} \subseteq \Sigma_{\mathcal{C}}^{\leftarrow \top}$ . The set of all  $\Sigma_{\mathcal{C}}^{\leftarrow \top}$ -reachable axioms is denoted by  $\mathcal{C}_{\Sigma}^{\leftarrow \top}$  and is called the  $\top$ -reachability module for  $\mathcal{C}$  over  $\Sigma$ .*

Given the appropriate mapping of axioms to hyperedges [16],  $\perp$ -Reachability can be shown to be equivalent to B-reachability in hypergraphs and  $\top$ -reachability to hypergraph F-reachability. The  $\perp$ -reachability module for a signature  $S$  is equivalent to the set of all B-hyperpaths for the set of nodes corresponding to  $S$  and the  $\top$ -reachability module equivalent to the set of all F-hyperpaths.

It is easy to show that  $\perp$ -reachability modules are equivalent to  $\perp$ -locality modules. However, by the definition of  $\top$ -reachability we observe that these are not equivalent to  $\top$ -locality modules.



*Example 2.* Let  $\mathcal{C}$  be a CBox such that  $\mathcal{C} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ , with  $\alpha_1 := A \sqsubseteq \exists r.D_1, \alpha_2 := B \sqsubseteq \geq 3r.D_2, \alpha_3 := \exists r.\top \sqsubseteq C, \alpha_4 := D_1 \sqsubseteq D_2$  and let  $\Sigma = \{C\}$ . Then  $\mathcal{C}_{\Sigma}^{\leftarrow \top} = \{\alpha_1, \alpha_2, \alpha_3\}$  but the  $\top$ -locality module for  $\mathcal{C}$  w.r.t.  $\Sigma$  is  $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ .

The difference stems from the fact that in  $\alpha_1$  and  $\alpha_2$  the  $\top$ -reachability of  $r$  does not ensure the  $\top$ -reachability of  $D_1$  and  $D_2$  respectively. This occurs because, given an axiom  $\alpha = \alpha_L \sqsubseteq \alpha_R$ ,  $\top$ -locality ensure that if  $\alpha$  is  $\top$ -local then so are all of the symbols in  $Sig(\alpha)$ , whereas  $\top$ -reachability is defined such that the  $\top$ -reachability of  $\alpha$  only guarantees that all symbols of  $\alpha_L$  and only some symbols of  $\alpha_R$  will be  $\top$ -reachable. Thus  $\top$ -reachability based modules are at most the size of  $\top$ -locality modules but in general could be substantially smaller. Similar to  $\perp \top^*$ -locality modules we note that reachability module extraction may also be alternated until a fixpoint is reached. These modules are denoted by  $\mathcal{C}_{\Sigma}^{\leftarrow \perp \top^*}$ .

Normalization plays an important role in the definition of reachability as the algorithm for determining  $\top$ -reachability of an axiom is different from the algorithm for determining  $\top$ -locality of an axiom. Not only does normalization simplify the definition of reachability considerably, it also allows us to determine exactly which symbols to exclude from our signature when adding new  $\top$ -reachable axioms. This can be seen in Example 2 where the symbol  $D_2$  is excluded when adding  $Sig(\alpha)$  to our signature. We note that a separate normalization phase is not strictly necessary, and that on-the-fly normalization can be done on an axiom during a reachability check. It is also possible to denormalize a normalized ontology by adding extra bookkeeping and labeling to the normalization process.

In order to investigate the module-theoretic properties of reachability modules, we follow a similar approach to Sattler et al. [18] and define inseparability different from that of conservative extensions. We say that  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are inseparable if their modules are equivalent, that is, a module extraction algorithm returns the same output for each of them. We define the following inseparability relations for reachability modules:

**Definition 11.** *Let  $\mathcal{C}_1$  and  $\mathcal{C}_2$  be CBoxes and  $\Sigma$  a signature. Then  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are:*

- $\Sigma - \top$  reachability inseparable, denoted by  $\mathcal{C}_1 \equiv_{\Sigma}^{\top} \mathcal{C}_2$ , if  $\mathcal{C}_1^{\leftarrow \top}_{\Sigma} = \mathcal{C}_2^{\leftarrow \top}_{\Sigma}$ ;
- $\Sigma - \perp$  reachability inseparable, denoted by  $\mathcal{C}_1 \equiv_{\Sigma}^{\perp} \mathcal{C}_2$ , if  $\mathcal{C}_1^{\leftarrow \perp}_{\Sigma} = \mathcal{C}_2^{\leftarrow \perp}_{\Sigma}$ ;
- $\Sigma - \perp \top^*$  reachability inseparable, denoted by  $\mathcal{C}_1 \equiv_{\Sigma}^{\perp \top^*} \mathcal{C}_2$ , if  $\mathcal{C}_1^{\leftarrow \perp \top^*}_{\Sigma} = \mathcal{C}_2^{\leftarrow \perp \top^*}_{\Sigma}$ .

Firstly we show that  $\top$ -reachability modules are subsumption inseparable. Concept inseparability follows as a special case of subsumption inseparability.

**Lemma 1.** *Let  $\mathcal{C}$  be a SRIQ CBox, and  $\Sigma \subseteq Sig(\mathcal{C})$  a signature. Let  $C, D$  be arbitrary SRIQ concept descriptions such that  $Sig(C) \cup Sig(D) \subseteq \Sigma$ . Then  $\mathcal{C} \models C \sqsubseteq D$  if and only if  $\mathcal{C}_{\Sigma}^{\leftarrow \top} \models C \sqsubseteq D$ .*

**Proof:** We have to prove two parts. First: If  $\mathcal{C}_{\Sigma}^{\leftarrow\top} \models C \sqsubseteq D$  then  $\mathcal{C} \models C \sqsubseteq D$ . This follows directly from the fact that  $\mathcal{C}_{\Sigma}^{\leftarrow\top} \subseteq \mathcal{C}$  and that  $\mathcal{SRIQ}$  is monotonic.

Conversely, we show that, if  $\mathcal{C} \models C \sqsubseteq D$  then  $\mathcal{C}_{\Sigma}^{\leftarrow\top} \models C \sqsubseteq D$ . Assume  $\mathcal{C} \models C \sqsubseteq D$  with  $\mathcal{I}_1$  a model for  $\mathcal{C}$ . Then there must exist an interpretation  $\mathcal{I}$  such that  $|\Delta^{\mathcal{I}}| \geq |\Delta^{\mathcal{I}_1}|$  and an individual  $w \in \Delta^{\mathcal{I}}$  such that  $\mathcal{I}$  is a model of  $\mathcal{C}_{\Sigma}^{\leftarrow\top}$  and  $w \in C^{\mathcal{I}} \setminus D^{\mathcal{I}}$ . Modify  $\mathcal{I}$  to  $\mathcal{I}'$  by setting  $x^{\mathcal{I}'} := \Delta^{\mathcal{I}}$  for all concept names  $x \in \text{Sig}(\mathcal{C}) \setminus \Sigma_{\mathcal{C}}^{\leftarrow\top}$ , and  $r^{\mathcal{I}'} := \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  for all roles names  $r \in \text{Sig}(\mathcal{C}) \setminus \Sigma_{\mathcal{C}}^{\leftarrow\top}$  and leaving everything else unchanged. We show that  $\mathcal{I}'$  is a model of  $\mathcal{C}_{\Sigma}^{\leftarrow\top}$ . For all  $\alpha := \alpha_L \sqsubseteq \alpha_R$ , with  $\alpha \in \mathcal{C}_{\Sigma}^{\leftarrow\top}$ , we have that:

- If  $\alpha_R$  is such that  $\text{Sig}(\alpha_R) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow\top}$  we have that  $(\alpha_R)^{\mathcal{I}} = (\alpha_R)^{\mathcal{I}'}$  since it does not change the interpretation of any symbols.
- If  $\alpha_R$  is an existential restriction of the form  $\exists r.A$  with  $y \in \text{Sig}(\alpha_R) \setminus \Sigma_{\mathcal{C}}^{\leftarrow\top}$ , then  $(y)^{\mathcal{I}'} = \Delta^{\mathcal{I}}$  or  $(y)^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  depending on whether  $y$  is a role or concept name. In both cases we have that  $(\alpha_R)^{\mathcal{I}} \subseteq (\alpha_R)^{\mathcal{I}'}$ .
- If  $\alpha_R$  is an at-least restriction of the form  $\geq nr.A$  with  $y \in \text{Sig}(\alpha_R) \setminus \Sigma_{\mathcal{C}}^{\leftarrow\top}$ , then  $(y)^{\mathcal{I}'} = \Delta^{\mathcal{I}}$  or  $(y)^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  depending on whether  $y$  is a role or concept name. In both cases we have that  $(\alpha_R)^{\mathcal{I}} \subseteq (\alpha_R)^{\mathcal{I}'}$ .
- If  $\alpha_R$  is of the form  $\exists R.\text{Self}$  with  $R \in \Sigma_{\mathcal{C}}^{\leftarrow\top}$  we have that  $(\alpha_R)^{\mathcal{I}} = (\alpha_R)^{\mathcal{I}'}$  since it does not change the interpretation of the symbol  $R$ .
- If  $\alpha$  is of the form  $\text{Dis}(R, S)$  then by definition it is always in  $\mathcal{C}_{\Sigma}^{\leftarrow\top}$ , thus  $R, S \in \Sigma_{\mathcal{C}}^{\leftarrow\top}$ . Therefore, the interpretation of alpha does not change.

In all cases  $(\alpha_L)^{\mathcal{I}} = (\alpha_L)^{\mathcal{I}'}$  since  $\alpha \in \mathcal{C}_{\Sigma}^{\leftarrow\top}$  and  $\text{Sig}(\alpha_L) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow\top}$  and thus  $(\alpha_L)^{\mathcal{I}'} \subseteq (\alpha_L)^{\mathcal{I}}$ . Thus,  $\mathcal{I}'$  is a model for  $\mathcal{C}_{\Sigma}^{\leftarrow\top}$ . Now for every  $\alpha = (\alpha_L \sqsubseteq \alpha_R) \in \mathcal{C} \setminus \mathcal{C}_{\Sigma}^{\leftarrow\top}$  we have:

- $\alpha_R$  is a concept name and  $\alpha_R^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ , or
- $\alpha_R$  is a role name and  $\alpha_R^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ , or
- $\alpha_R$  is a disjunction of the form  $A_1 \sqcup \dots \sqcup A_n$  with at least one  $A_i \notin \Sigma_{\mathcal{C}}^{\leftarrow\top}$ , thus  $A_i^{\mathcal{I}'} = \Delta^{\mathcal{I}}$  and  $\alpha_R^{\mathcal{I}'} = A_1^{\mathcal{I}} \cup \dots \cup \Delta^{\mathcal{I}} \cup \dots \cup A_n^{\mathcal{I}} = \Delta^{\mathcal{I}}$ , or
- $\alpha_R$  is an existential restriction  $\exists r.A_1$ , thus  $r^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  and  $A_1^{\mathcal{I}'} = \Delta^{\mathcal{I}}$  so that  $(\exists r.A_1)^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ , or
- $\alpha_R$  is  $\exists r.\text{Self}$ , thus  $r^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  so that  $(\exists r.\text{Self})^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ , or
- $\alpha_R$  is an atleast restriction  $\geq nr.A_2$ , thus  $r^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ ,  $A_2^{\mathcal{I}'} = \Delta^{\mathcal{I}}$  and  $|\Delta^{\mathcal{I}}| \geq n$  so that  $(\geq nr.A_2)^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ . This follows from the fact that  $|\Delta^{\mathcal{I}}| \geq |\Delta^{\mathcal{I}_1}|$  and for any concept description  $\geq nr.A$ ,  $|\Delta^{\mathcal{I}}| \geq |(r.A)^{\mathcal{I}}| \geq n$  for it to be satisfiable.

Since for all cases  $\alpha_L^{\mathcal{I}'} \subseteq \alpha_R^{\mathcal{I}'}$ , we conclude that  $\mathcal{I}'$  is a model for  $\mathcal{C}$ . But  $\mathcal{I}$  and  $\mathcal{I}'$  correspond on all symbols  $y \in (\text{Sig}(D) \cup \text{Sig}(C)) \subseteq \Sigma \subseteq \Sigma_{\mathcal{C}}^{\leftarrow\top}$  and therefore  $D^{\mathcal{I}'} = D^{\mathcal{I}}$  and  $C^{\mathcal{I}'} = C^{\mathcal{I}}$ . Now since  $C^{\mathcal{I}} = C^{\mathcal{I}'}$  and  $w \in C^{\mathcal{I}}$  we have that  $w \in C^{\mathcal{I}'} \setminus D^{\mathcal{I}'}$  and hence  $\mathcal{C} \not\models C \sqsubseteq D$ , contradicting the assumption.  $\square$

**Corollary 1.** *Let  $\mathcal{C}$  be a normalized  $\mathcal{SRIQ}$  CBox,  $\Sigma \subseteq \text{Sig}(\mathcal{C})$  a signature and  $S$  an inseparability relation from Definitions 3 and 4. Then  $\mathcal{C}_{\Sigma}^{\leftarrow\top} \equiv_{\Sigma}^S \mathcal{C}$ .  $\mathcal{C}_{\Sigma}^{\leftarrow\top}$  is therefore a  $S_{\Sigma}$ -module of  $\mathcal{C}$ .*

We show by way of counter example that  $\mathcal{C}_{\Sigma}^{\leftarrow \top}$  is not a self-contained or depleting  $S_{\Sigma}$  module of  $\mathcal{C}$  when  $\Sigma_{\mathcal{C}}^{\leftarrow \top} \neq \text{Sig}(\mathcal{C}_{\Sigma}^{\leftarrow \top})$ .

*Example 3.* Let  $\mathcal{C}$  be a CBox such that  $\mathcal{C} = \{\alpha_1 = A \sqsubseteq \exists r.D_1, \alpha_2 = B \sqsubseteq \geq nr.D_2, \alpha_3 = \exists r.\top \sqsubseteq C, \alpha_4 = D_1 \sqsubseteq D_2\}$ , and let  $\Sigma = \{C\}$ . Then  $\mathcal{C}_{\Sigma}^{\leftarrow \top} = \{\alpha_1, \alpha_2, \alpha_3\}$ ,  $\delta = \Sigma \cup \text{Sig}(\mathcal{C}_{\Sigma}^{\leftarrow \top}) = \{A, B, C, r, D_1, D_2\} \neq \Sigma_{\mathcal{C}}^{\leftarrow \top}$ . But  $\mathcal{C} \models D_1 \sqsubseteq D_2$  and  $\mathcal{C}_{\Sigma}^{\leftarrow \top} \not\models D_1 \sqsubseteq D_2$ . Therefore  $\mathcal{C}_{\Sigma}^{\leftarrow \top}$  is not a self-contained  $c_{\Sigma}$ -module of  $\mathcal{C}$ . Similarly,  $\mathcal{C} \setminus \mathcal{C}_{\Sigma}^{\leftarrow \top} \models \alpha_4 \neq \emptyset$  with  $\Sigma = D_1, D_2$  and  $D_1, D_2 \in \delta$ . Therefore,  $\mathcal{C}_{\Sigma}^{\leftarrow \top}$  is not a depleting  $c_{\Sigma}$ -module of  $\mathcal{C}$ .

Before investigating the robustness properties of reachability modules we introduce some lemmas to aid us in the proofs that follow.

**Lemma 2.** *Let  $\alpha$  be an axiom,  $\Sigma$  and  $\Sigma'$  be signatures,  $x \in \{\top, \perp\}$  and  $\mathcal{C}$  a SRIQ CBox. Then:*

1. *If  $\Sigma \subseteq \Sigma'$  and  $\alpha$  is not  $\Sigma'_{\mathcal{C}}^{\leftarrow x}$  reachable, then  $\alpha$  is not  $\Sigma_{\mathcal{C}}^{\leftarrow x}$  reachable.*
2. *If  $\Sigma' \cap \text{Sig}(\alpha) \subseteq \Sigma$  and  $\alpha$  is not  $\Sigma$  reachable then  $\alpha$  is not  $\Sigma'$  reachable.*

**Proof:**

1. By the inductive definition of  $x$ -reachability if  $\Sigma \subseteq \Sigma'$  then  $\Sigma_{\mathcal{C}}^{\leftarrow x} \subseteq \Sigma'_{\mathcal{C}}^{\leftarrow x}$ . Thus if  $\alpha$  is not  $\Sigma'_{\mathcal{C}}^{\leftarrow x}$  reachable it can also not be  $\Sigma_{\mathcal{C}}^{\leftarrow x}$ -reachable.
2. Assume that  $\alpha$  is not  $\Sigma$  reachable but it is  $\Sigma'$  reachable. Then there is at least one symbol  $y \in \text{Sig}(\alpha)$  such that  $y \notin \Sigma$  and  $\alpha$  is  $\Sigma \cup \{y\}$  reachable.  $\alpha$  is  $\Sigma'$  reachable so it must be the case that  $y \in \Sigma'$ . But this contradicts our assumption that  $\Sigma' \cap \text{Sig}(\alpha) \subseteq \Sigma$ . Thus,  $\alpha$  is not  $\Sigma'$  reachable.

**Lemma 3.** *Let  $\alpha$  be an axiom,  $\Sigma$  and  $\Sigma'$  be signatures,  $x \in \{\top, \perp\}$  and  $\mathcal{C}, \mathcal{C}'$  SRIQ CBoxes. Then:*

1. *Given  $\mathcal{C}_1 = \mathcal{C}_{\Sigma'}^{\leftarrow x}$ , if  $\Sigma \subseteq \Sigma'$  then  $\mathcal{C}_{\Sigma}^{\leftarrow x} = \mathcal{C}_1^{\leftarrow x}$ . In particular  $\mathcal{C}_{\Sigma}^{\leftarrow x} \subseteq \mathcal{C}_{\Sigma'}^{\leftarrow x}$ .*
2. *If  $\Sigma' \cap \text{Sig}(\mathcal{C}) \subseteq \Sigma$ , then  $\mathcal{C}_{\Sigma'}^{\leftarrow x} \subseteq \mathcal{C}_{\Sigma}^{\leftarrow x}$ .*
3. *If  $\mathcal{C} \subseteq \mathcal{C}'$ , then  $\mathcal{C}_{\Sigma}^{\leftarrow x} \subseteq \mathcal{C}'_{\Sigma}^{\leftarrow x}$ .*

**Proof:**

1. Assume that there is some axiom  $\alpha \in \mathcal{C}_{\Sigma}^{\leftarrow x}$  such that  $\alpha \notin \mathcal{C}_{\Sigma'}^{\leftarrow x}$ . Therefore, we have that  $\alpha$  is not  $\Sigma'_{\mathcal{C}}^{\leftarrow x}$  reachable but that it is  $\Sigma_{\mathcal{C}}^{\leftarrow x}$  reachable. But this is a contradiction by Lemma 2.1 since  $\Sigma \subseteq \Sigma'$ . Thus,  $\mathcal{C}_{\Sigma}^{\leftarrow x} \subseteq \mathcal{C}_{\Sigma'}^{\leftarrow x}$ . A similar argument is used to show that  $\mathcal{C}_{\Sigma}^{\leftarrow x} \subseteq \mathcal{C}_1^{\leftarrow x}$  and  $\mathcal{C}_1^{\leftarrow x} \subseteq \mathcal{C}_{\Sigma}^{\leftarrow x}$ .
2. For every  $\alpha \in \mathcal{C}$  we have that  $\Sigma' \cap \text{Sig}(\alpha) \subseteq \Sigma$ . Therefore, from Lemma 2.2 we have that whenever  $\alpha \in \mathcal{C}$  is not  $\Sigma$  reachable it is also not  $\Sigma'$  reachable and we have that  $\mathcal{C}_{\Sigma'}^{\leftarrow x}$  contains at most all those axioms in  $\mathcal{C}_{\Sigma}^{\leftarrow x}$ . Thus,  $\mathcal{C}_{\Sigma'}^{\leftarrow x} \subseteq \mathcal{C}_{\Sigma}^{\leftarrow x}$ .
3. Let  $\delta = \Sigma_{\mathcal{C}}^{\leftarrow x}$ ,  $\delta' = \Sigma_{\mathcal{C}_1}^{\leftarrow x}$  and  $\alpha \in (\mathcal{C} \cap \mathcal{C}_1)$ . Assume  $\alpha$  is  $\delta$  reachable but not  $\delta'$  reachable. Since  $\mathcal{C} \subseteq \mathcal{C}_1$  and  $\text{Sig}(\mathcal{C}) \subseteq \text{Sig}(\mathcal{C}_1)$  we have by the inductive definition of  $x$  reachability that  $\delta \subseteq \delta'$ . But by Lemma 2.1 we have that whenever  $\alpha$  is not  $\delta'$  reachable then it is also not  $\delta$  reachable. Therefore,  $\mathcal{C}_{\Sigma}^{\leftarrow x}$  contains at most all those axioms in  $\mathcal{C}_1^{\leftarrow x}$ . Thus,  $\mathcal{C}_{\Sigma}^{\leftarrow x} \subseteq \mathcal{C}_1^{\leftarrow x}$ .

**Lemma 4.** *Let  $\Sigma$  be a signature,  $\mathcal{C}_1$  and  $\mathcal{C}_2$  be *SRIQ* CBoxes with  $\text{Sig}(\mathcal{C}_1) \cap \text{Sig}(\mathcal{C}_2) \subseteq \Sigma$  and  $x \in \{\top, \perp\}$ . Then  $(\mathcal{C}_1 \cup \mathcal{C}_2)_{\Sigma}^{\leftarrow x} = \mathcal{C}_{1\Sigma}^{\leftarrow x} \cup \mathcal{C}_{2\Sigma}^{\leftarrow x}$ .*

**Proof:** Let  $\mathcal{M} = (\mathcal{C}_1 \cup \mathcal{C}_2)_{\Sigma}^{\leftarrow x}$ ,  $\mathcal{M}_1 = \mathcal{C}_{1\Sigma}^{\leftarrow x}$ ,  $\mathcal{M}_2 = \mathcal{C}_{2\Sigma}^{\leftarrow x}$ . Now  $\mathcal{C}_1 \subseteq \mathcal{C}_1 \cup \mathcal{C}_2$  thus by Lemma 3.3 we have that  $\mathcal{M}_1 \subseteq \mathcal{M}$ . Similarly  $\mathcal{M}_2 \subseteq \mathcal{M}$  and thus  $\mathcal{M}_1 \cup \mathcal{M}_2 \subseteq \mathcal{M} \cup \mathcal{M}$  which gives us  $\mathcal{M}_1 \cup \mathcal{M}_2 \subseteq \mathcal{M}$ . Let  $\Sigma' = \Sigma \cup \Sigma_{\mathcal{C}_1}^{\leftarrow x} \cup \Sigma_{\mathcal{C}_2}^{\leftarrow x}$ . To show that  $\mathcal{M} \subseteq \mathcal{M}_1 \cup \mathcal{M}_2$  we note that, when extracting these modules, the order in which axioms are extracted are irrelevant. We therefore assume that any algorithm first extracts axioms in  $\mathcal{M}_1 \cup \mathcal{M}_2$  then tests all other axioms for  $\Sigma'_{\mathcal{C}_1 \cup \mathcal{C}_2}$ -reachability. Consider any axiom  $\alpha \in (\mathcal{C}_1 \cup \mathcal{C}_2) \setminus (\mathcal{M}_1 \cup \mathcal{M}_2)$ . If  $\alpha \in \mathcal{C}_1$  then  $\alpha \in \mathcal{C}_1 \setminus \mathcal{M}_1$  and  $\alpha$  is not  $\Sigma_{\mathcal{C}_1}^{\leftarrow x} \cup \Sigma$  reachable. Now precondition  $\text{Sig}(\mathcal{C}_2) \cap \text{Sig}(\mathcal{C}_1) \subseteq \Sigma$  implies  $\Sigma_{\mathcal{C}_2}^{\leftarrow x} \cap \text{Sig}(\alpha) \subseteq \Sigma$ , taken that  $\alpha$  is not  $\Sigma_{\mathcal{C}_1}^{\leftarrow x} \cup \Sigma$  reachable we manipulate this statement to derive  $(\Sigma \cup \Sigma_{\mathcal{C}_2}^{\leftarrow x} \cup \Sigma_{\mathcal{C}_1}^{\leftarrow x}) \cap \text{Sig}(\alpha) \subseteq \Sigma \cup \Sigma_{\mathcal{C}_1}^{\leftarrow x}$ . Thus by Lemma 2.2 we have that  $\alpha$  is not  $\Sigma \cup \Sigma_{\mathcal{C}_2}^{\leftarrow x} \cup \Sigma_{\mathcal{C}_1}^{\leftarrow x}$  reachable. The case where  $\alpha \in \mathcal{C}_2$  is treated analogously.  $\square$

**Proposition 1.** *For  $x \in \{\top, \perp\}$ ,  $x$ -reachability is robust under replacement.*

**Proposition 2.** *For  $x \in \{\top, \perp\}$ ,  $x$ -reachability is robust under vocabulary extensions.*

**Proposition 3.** *For  $x \in \{\top, \perp\}$ ,  $x$ -reachability is robust under vocabulary restrictions.*

**Proposition 4.** *For  $x \in \{\top, \perp\}$ ,  $x$ -reachability is robust under joins.*

The proofs to show that reachability modules including  $\mathcal{C}_{\Sigma}^{\leftarrow \perp \top^*}$  modules share all the robustness properties of locality modules follow from the above lemmas and follow the proofs for locality modules by Sattler, et al. [18].

Reachability modules therefore share all the robustness properties listed. However, we have seen that these modules are neither depleting nor self-contained modules. Amongst other things, the depleting and self-contained nature of modules are utilised in order to find all justifications for an entailment [8].

**Definition 12.** *Let  $\mathcal{C}$  be a *SRIQ* CBox and  $\mathcal{M} \subseteq \mathcal{C}$ .  $\mathcal{M}$  is a justification for  $\mathcal{C} \models C \sqsubseteq D$  if  $\mathcal{M} \models C \sqsubseteq D$  and there exists no  $\mathcal{M}_1 \subset \mathcal{M}$  such that  $\mathcal{M}_1 \models C \sqsubseteq D$ .*

We show that although our modules do not share these properties they do contain all justifications for a given signature.

**Theorem 2.** *Let  $\mathcal{C}$  be a normalized *SRIQ* CBox and  $\Sigma$  a signature such that  $\Sigma \subseteq \text{Sig}(\mathcal{C})$ . Then for arbitrary concept descriptions  $C, D$ , such that  $\mathcal{C} \models C \sqsubseteq D$  and  $\text{Sig}(C) \cup \text{Sig}(D) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow \top}$  we have that  $\mathcal{C}_{\Sigma}^{\leftarrow \top}$  contains all justifications for  $\mathcal{C} \models C \sqsubseteq D$ .*

**Proof:** Assume that  $\mathcal{C} \models C \sqsubseteq D$  for some  $\text{Sig}(C) \cup \text{Sig}(D) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow \top}$ , but there is a justification  $\mathcal{M}$  for  $\mathcal{C} \models C \sqsubseteq D$  that is not contained in  $\mathcal{C}_{\Sigma}^{\leftarrow \top}$ . If  $C \sqsubseteq D$  is a tautology then  $\mathcal{M}$  must be empty with  $\mathcal{M} \subseteq \mathcal{C}_{\Sigma}^{\leftarrow \top}$ . Thus, we assume that  $C \sqsubseteq D$

is not a tautology. Since  $\mathcal{M} \not\subseteq \mathcal{C}_{\Sigma}^{\leftarrow\top}$ , there must be an axiom  $\alpha \in \mathcal{M} \setminus \mathcal{C}_{\Sigma}^{\leftarrow\top}$ . Define  $\mathcal{M}_1 := \mathcal{M} \cap \mathcal{C}_{\Sigma}^{\leftarrow\top}$ .  $\mathcal{M}_1$  is a strict subset of  $\mathcal{M}$  since  $\alpha \notin \mathcal{M}_1$ . There are two cases, either  $\mathcal{M}_1 = \emptyset$  or it contains at least one axiom.

In the case where  $\mathcal{M}_1 = \emptyset$ , define  $\mathcal{C}_1 = \mathcal{C} \setminus \mathcal{C}_{\Sigma}^{\leftarrow\top}$  with  $\mathcal{M} \subseteq \mathcal{C}_1$ . Now since  $\mathcal{M} \models C \sqsubseteq D$  we have by monotonicity that  $\mathcal{C}_1 \models C \sqsubseteq D$ . Since  $\mathcal{C}_1 \subseteq \mathcal{C}$  we have by Lemma 3.3 that  $\mathcal{C}_1^{\leftarrow\top} \subseteq \mathcal{C}_{\Sigma}^{\leftarrow\top}$  and thus that  $\mathcal{C}_1^{\leftarrow\top} = \emptyset$ . But by Lemma 1 we have that  $\mathcal{C}_1^{\leftarrow\top} \models C \sqsubseteq D$  if, and only if,  $\mathcal{C}_1 \models C \sqsubseteq D$ . Since  $C \sqsubseteq D$  is not a tautology we have that  $\mathcal{C}_1^{\leftarrow\top} \not\models C \sqsubseteq D$  and thus that  $\mathcal{M} \not\models C \sqsubseteq D$ .

In the case where  $\mathcal{M}_1 \neq \emptyset$  we claim that  $\mathcal{M}_1 \models C \sqsubseteq D$ , which contradicts the fact that  $\mathcal{M}$  is a justification for  $\mathcal{C} \models C \sqsubseteq D$ .

We use proof by contraposition to show this. Assume that  $\mathcal{M}_1 \not\models C \sqsubseteq D$ , i.e., there is a model  $\mathcal{I}_1$  of  $\mathcal{M}_1$  such that  $C^{\mathcal{I}_1} \not\subseteq D^{\mathcal{I}_1}$ . We modify  $\mathcal{I}_1$  to  $\mathcal{I}$  by setting  $y^{\mathcal{I}} := \Delta^{\mathcal{I}_1}$  for all concept names  $y \in \text{Sig}(\mathcal{C}) \setminus \Sigma_{\mathcal{C}}^{\leftarrow\top}$ , and  $r^{\mathcal{I}} := \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_1}$  for all roles names  $r \in \text{Sig}(\mathcal{C}) \setminus \Sigma_{\mathcal{C}}^{\leftarrow\top}$ . We have  $D^{\mathcal{I}} = D^{\mathcal{I}_1}$  since  $\text{Sig}(D) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow\top}$ , and  $C^{\mathcal{I}} = C^{\mathcal{I}_1}$  since  $\text{Sig}(C) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow\top}$ . It follows that  $C^{\mathcal{I}} \not\subseteq D^{\mathcal{I}}$ . It remains to be shown that  $\mathcal{I}$  is indeed a model of  $\mathcal{M}$ , and therefore satisfies all axioms  $\beta = (\beta_L \sqsubseteq \beta_R)$  in  $\mathcal{M}$ , including  $\alpha$ . If  $\beta = \text{Dis}(R_r, R_2)$  then by definition  $\text{Sig}(\beta) \subseteq \Sigma_{\mathcal{C}}^{\leftarrow\top}$  so that  $(\beta)^{\mathcal{I}} = (\beta)^{\mathcal{I}_1}$ . Otherwise there are two possibilities:

- $\beta \in \mathcal{M}_1$ . Since  $\mathcal{M}_1 \subseteq \mathcal{C}_{\Sigma}^{\leftarrow\top}$ , all symbols in  $\text{Sig}(\beta_L)$  are also in  $\Sigma_{\mathcal{C}}^{\leftarrow\top}$  and possibly some symbols of  $\text{Sig}(\beta_R)$  may not be in  $\Sigma_{\mathcal{C}}^{\leftarrow\top}$ . Consequently,  $\mathcal{I}_1$  and  $\mathcal{I}$  coincide on the names occurring in  $\beta_L$  and since  $\mathcal{I}_1$  is a model of  $\mathcal{M}_1$ , we have that  $(\beta_L)^{\mathcal{I}} = (\beta_L)^{\mathcal{I}_1}$  and  $(\beta_R)^{\mathcal{I}_1} \subseteq (\beta_R)^{\mathcal{I}}$ . Therefore  $(\beta_L)^{\mathcal{I}} \subseteq (\beta_R)^{\mathcal{I}}$ .
- $\beta \notin \mathcal{M}_1$ . Since  $\beta \in \mathcal{M}$ , we have that  $\beta \notin \mathcal{C}_{\Sigma}^{\leftarrow\top}$ , and hence  $\beta$  is not  $\Sigma_{\mathcal{C}}^{\leftarrow\top}$ -reachable. Thus,
  - $\beta_R$  is a concept name and  $\beta_R^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ , or
  - $\beta_R$  is a role name and  $\beta_R^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ , or
  - $\beta_R$  is a disjunction of the form  $A_1 \sqcup \dots \sqcup A_n$  with at least one  $A_i \notin \Sigma_{\mathcal{C}}^{\leftarrow\top}$ , thus  $A_i^{\mathcal{I}'} = \Delta^{\mathcal{I}}$  and  $\beta_R^{\mathcal{I}'} = A_1^{\mathcal{I}} \cup \dots \cup \Delta^{\mathcal{I}} \cup \dots \cup A_n^{\mathcal{I}} = \Delta^{\mathcal{I}}$ , or
  - $\beta_R$  is an existential restriction  $\exists r.A_1$ , thus  $r^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  and  $A_1^{\mathcal{I}'} = \Delta^{\mathcal{I}}$  so that  $(\exists r.A_1)^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ , or
  - $\beta_R$  is  $\exists r.\text{Self}$ , thus  $r^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  so that  $(\exists r.\text{Self})^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ , or
  - $\beta_R$  is an atleast restriction  $\geq nr.A_2$ , thus  $r^{\mathcal{I}'} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ ,  $A_2^{\mathcal{I}'} = \Delta^{\mathcal{I}}$  and  $|\Delta^{\mathcal{I}}| \geq n$  so that  $(\geq nr.A_2)^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ . This follows from the fact that for any concept description  $\geq nr.A$ ,  $|\Delta^{\mathcal{I}}| \geq |(r.A)^{\mathcal{I}}| \geq n$  for it to be satisfiable.

By definition of  $\mathcal{I}$ ,  $(\beta_R)^{\mathcal{I}} = \Delta^{\mathcal{I}_1}$ . Hence  $(\beta_L)^{\mathcal{I}} \subseteq (\beta_R)^{\mathcal{I}}$ .

Therefore  $\mathcal{I}$  is a model for  $\mathcal{M}$ . But since  $C^{\mathcal{I}} \not\subseteq D^{\mathcal{I}}$  we have that  $\mathcal{M} \not\models C \sqsubseteq D$  proving the contrapositive.  $\square$

## 5 Empirical Evaluation

From Example 2 we see that reachability modules have the potential of being smaller than locality modules. In this section we show the results of tests conducted to determine the extent of the difference in size between reachability and

locality modules across a range of real world ontologies. The criteria used to select the target ontologies were size and expressivity. In terms of size we tried to find ontologies that range from a few thousand to tens of thousands of CBox axioms. In terms of expressivity we chose ontologies that range from the relatively inexpressive DL  $\mathcal{EL}$  up to and including  $\mathcal{SRIQ}$ . For ontologies containing nominals we simple removed all axioms containing nominals from the test ontology. In Table 3 we provide a non-exhaustive list of DL metrics for each of the ontologies<sup>1</sup> in the test set.

**Table 3.** DL Metrics

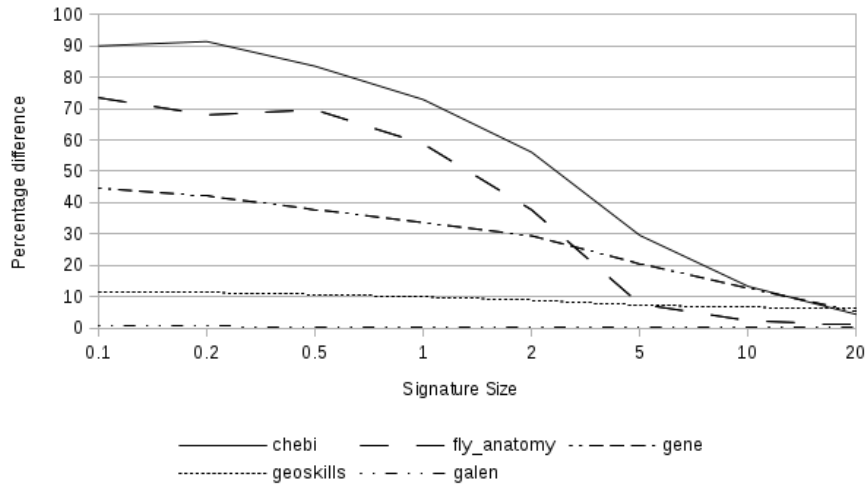
	Chebi	Fly_Anatomy	Gene	GeoSkills	Galen	cton	so-xp	Software
Expressivity	$\mathcal{EL}^{++}$	$\mathcal{EL}^{++}$	$\mathcal{EL}^{++}$	$ALCHOIN(D)$	$AL\mathcal{E}HF^+$	$SHF$	$SHI$	$ALCHIQ(D)$
Axioms	34387	10471	42656	14861	4735	33203	1943	3347
Concepts	19360	6222	26225	603	2748	17033	1660	735
Roles	8	2	4	23	413	43	22	15
$C \sqsubseteq D$	34387	10467	42650	686	3237	33062	1709	2077
$C \equiv D$	0	2	0	6	0	86	198	7
$C \sqcap D \sqsubseteq \perp$	0	0	2	19	0	8	21	0
$Trans(R)$	0	2	1	0	26	18	5	0
$R \sqsubseteq S$	0	0	2	4	416	25	6	1
$R^-$	0	0	0	1	207	0	0	3
$Ran(R)$	0	0	0	15	0	0	0	4
$Dom(R)$	0	0	0	16	0	0	0	3
$Sym(R)$	0	0	0	1	0	0	4	0

Test were structured in such a way that we could determine the difference in module sizes across a range of different input signature sizes. For each of the test ontologies  $\mathcal{O}_i$  we chose a random signature as a percentage of  $Sig(\mathcal{O}_i)$ . The input signature size was divided into eight groups namely 0.1%, 0.2%, 0.5%, 1.0%, 2.0%, 5.0%, 10.0% and 20.0% of  $Sig(\mathcal{O}_i)$ . For each of these input sizes we extracted one thousand  $\perp \top^*$ -reachability and locality modules, each module based on a random selection of symbols from  $Sig(\mathcal{O}_i)$  to act as input signature  $\mathcal{S}$ . The average difference in size between reachability and locality modules were then calculated by the formula  $Avg((Local_j(\mathcal{S}_j) - Reach_j(\mathcal{S}_j)) * 100 / (Local_j(\mathcal{S}_j)))$  for  $1 \leq j \leq 1000$ .

Figure 1 represents the reduction in size of reachability modules versus locality modules. The  $x$ -axis represents the signature size whereas the  $y$ -axis repre-

<sup>1</sup> Obtained from the TONES repository 15 July 2013  
(<http://owl.cs.manchester.ac.uk/repository/browser>).

**Fig. 1.** Reachability v.s. Locality Modules



sents percentage reduction of reachability modules over that of locality modules. From this graph we see that there is a drastic difference between the results for different ontologies. For relatively small signature sizes in the **Chebi** ontology reachability modules can be up to 90% smaller than locality modules for the same input signature, whereas for the **galen** ontology there is less than 1% difference. For the **so-xp**, **cton** and **software** ontologies, not listed here, the results are very similar to that of **galen**.

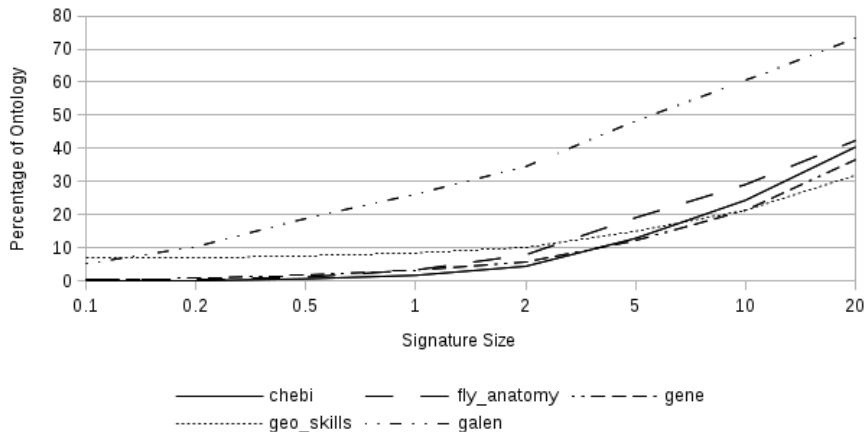
Figure 2 represents the ratio between the size of the reachability module v.s. the size of the whole ontology. The  $x$ -axes represents the signature size and the  $y$ -axis represents the reachability module size as a percentage of the ontology size.

From the results we see that reachability modules are potentially smaller than locality modules. The drastic difference in the results further demonstrate that reachability modules may be of better use where the input signature is relatively small. For the current set of results we have not attempted to deduce the reasons why certain ontologies perform better than others.

## 6 Conclusion

We have investigated the module-theoretic properties of reachability modules for *SRIQ* CBoxes. Reachability modules differ from syntactic locality modules in that they are not self-contained or depleting. One application of the self-contained and depleting nature of locality modules is to find all justifications for entailments. However, in terms of finding justifications, by showing that

**Fig. 2.** Reachability module v.s. Ontology



reachability modules do preserve all justifications for entailments, we have shown that these properties are sufficient but that they are not necessary.

We did an empirical evaluation into the size difference between locality and reachability modules. We extracted a random sample of 1000 modules from each of the ontologies listed in Table 3. Reachability modules were between 0% and 90% smaller than locality modules with a relatively small input signature. This difference diminishes when the signature size reaches over 20% of the signature size of the ontology.

Our focus for future research is to extend these results to *SRIOQ* and to investigate the relationship between other hypergraph based problems and DL reasoning problems more closely.

**Acknowledgments** This work is based upon research supported in part by the National Research Foundation of South Africa (UID 85482, IFR2011032700018).

## References

1. Antoniou, G., Kehagias, A.: A note on the refinement of ontologies. *International Journal of Intelligent Systems* 15, 623–632 (2000)
2. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): *The description logic handbook: theory, implementation, and applications*. Cambridge University Press, New York, NY, USA (2003)
3. Cuenca Grau, B., Halaschek-Wiener, C., Kazakov, Y., Suntisrivaraporn, B.: *Incremental Classification of Description Logic Ontologies*. Tech. rep. (2012)
4. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Just the right amount: extracting modules from ontologies. In: Williamson, C., Zurko, M. (eds.) *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. pp. 717–726. ACM, New York, NY, USA (2007)



5. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Modular Reuse of Ontologies: Theory and Practice. *Journal of Artificial Intelligence Research (JAIR)* 31, 273–318 (2008)
6. Del Vescovo, C., Parsia, B., Sattler, U., Schneider, T.: The modular structure of an ontology: atomic decomposition and module count. In: O. Kutz, T.S. (ed.) *Proc. of WoMO-11. Frontiers in AI and Appl.*, vol. 230, pp. 25–39. IOS Press (2011)
7. Ghilardi, S., Lutz, C., Wolter, F.: Did I damage my ontology? a case for conservative extensions in description logics. In: Doherty, P., Mylopoulos, J., Welty, C. (eds.) *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning (KR'06)*. pp. 187–197. AAAI Press (2006)
8. Horridge, M., Parsia, B., Sattler, U.: Laconic and precise justifications in owl. In: *In Proc. of ISWC-08, volume 5318 of LNCS*. pp. 323–338 (2008)
9. Horrocks, I., Kutz, O., Sattler, U.: The irresistible *SRIQ*. In: *In Proc. of OWL: Experiences and Directions* (2005)
10. Konev, B., Lutz, C., Walther, D., Wolter, F.: Formal properties of modularisation. In: Stuckenschmidt, H., Parent, C., Spaccapietra, S. (eds.) *Modular Ontologies*, chap. 2, pp. 25–66. Springer-Verlag, Berlin, Heidelberg (2009)
11. Kontchakov, R., Pulina, L., Sattler, U., 0002, T.S., Selmer, P., Wolter, F., Zakharyashev, M.: Minimal module extraction from dl-lite ontologies using qbf solvers. In: Boutilier, C. (ed.) *IJCAI*. pp. 836–841 (2009)
12. Kontchakov, R., Wolter, F., Zakharyashev, M.: Logic-based ontology comparison and module extraction, with an application to dl-lite. *Artificial Intelligence* 174(15), 1093–1141 (2010)
13. Nguyen, S., Pretolani, D., Markenzon, L.: On some path problems on oriented hypergraphs. *Theoretical Informatics and Applications* 32(1-2-3), 1–20 (1998)
14. Nortjé, R.: Module extraction for inexpressive description logics. Master’s thesis, University of South Africa (2011)
15. Nortjé, R., Britz, K., Meyer, T.: Bidirectional reachability-based modules. In: *Proceedings of the 2011 International Workshop on Description Logics (DL2011)*. CEUR Workshop Proceedings, CEUR-WS (2011), <http://ceur-ws.org>
16. Nortjé, R., Britz, K., Meyer, T.: A normal form for hypergraph-based module extraction for *SROIQ*. In: Gerber, A., Taylor, K., Meyer, T., Orgun, M. (eds.) *Australasian Ontology Workshop 2009 (AOW 2012)*. Ceur-ws, vol. 969, pp. 40–51. CEUR, Melbourne, Australia (2012), <http://ceur-ws.org/Vol-969/proceedings.pdf>
17. Nortjé, R., Britz, K., Meyer, T.: Module-theoretic properties of reachability modules for *SRIQ*. In: *Description Logic Workshop 2013 (DL2013)*. CEUR-WS, CEUR, Ulm, Germany (2013)
18. Sattler, U., Schneider, T., Zakharyashev, M.: Which kind of module should I extract? In: Grau, B.C., Horrocks, I., Motik, B., Sattler, U. (eds.) *Description Logics. CEUR Workshop Proceedings*, vol. 477. CEUR-WS.org (2009)
19. Suntisrivaraporn, B.: Polynomial-Time Reasoning Support for Design and Maintenance of Large-Scale Biomedical Ontologies. Ph.D. thesis, Technical University of Dresden (2009)
20. Thakur, M., Tripathi, R.: Complexity of Linear Connectivity Problems in Directed Hypergraphs. *Linear Connectivity Conference* pp. 1–12 (2001)