

# Boundary-based Module Extraction in $\mathcal{EL}^{++}$ Ontologies

Jun Fang<sup>1</sup>, Jie Bao<sup>2</sup> and Lei Guo<sup>1</sup>

<sup>1</sup> Department of Automation, Northwestern Polytechnical University, China.  
{junfang, lguo}@nwpu.edu.cn

<sup>2</sup> Department of Computer Science, Rensselaer Polytechnic Institute, USA.  
baojie@cs.rpi.edu

**Abstract.** Modularization is a promising technique to meet the scalability challenge in reasoning with very large ontologies. In this work, we introduce a novel boundary-based modular extraction method for ontologies in  $\mathcal{EL}^{++}$  description logics. The proposed method is capable of identifying relevant axioms in an ontology based on the notion of boundaries of symbols, with respect to a given reasoning task. We present the theoretical foundation and a practical algorithm for computing boundary-based modules. As most work of boundary-based module extraction can be done offline, the modularization results can be applied in improving reasoning performance. The proposed algorithm is implemented for the DL  $\mathcal{EL}^{++}$ . Experimental results on real-world ontologies show that boundary-based modules generated by our method are very close to the optimal result.

## 1 Introduction

Modularization is an important technique that may facilitate reasoning with very large ontologies. With the identification of the modular structure of an ontology, a reasoning task may be carried out more efficiently by pinpointing the relevant modules, or by dividing the overall reasoning task into smaller components that can be computed against modules [1, 2, 3]. In addition, modules can also be used in other ontology reasoning tasks such as incremental reasoning [4] and analogical reasoning. To achieve high accuracy and efficiency of inference, a modularization-based reasoning approach needs to meet several critical requirements. In particular, we need to ensure: (i) *Exactness*, which means that the answer of a reasoning task performed in the modular way and the answer to the same task that is performed in the conventional way on the whole ontology are always identical; (ii) *Compactness*, which means that a reasoning task should involve the minimal set of modules or axioms that are necessary for performing the reasoning task. The goal of this work is to investigate a practical method of module extraction that can satisfy those requirements and is suitable to improve reasoning scalability.

There is a growing attention on modular ontologies. Please see [5] for a survey of modular ontologies in general. Most relevant work to our study is the work on ontology modularization using structural [6, 7, 8] or logic-based approaches [9, 10, 11]. However, modules extracted by using structural methods in general may not be exact in query answering. Logic-based approaches provide ways to extract modules with exactness

guarantee, e.g., by signature-based module extraction. In this method, given a set of symbols (i.e., a signature), a fragment can be extracted from an ontology that can answer any reasoning problems for that signature in the exact manner. However, the modules discovered by this approach are in general not compact [12, 13].

In this work, we present a novel axiom-based approach for module extraction that behaves well for both exactness and compactness. Signature-based modularization is targeted at extracting a module that can preserve all knowledge about the signature in question from the entire ontology. This may result that axioms that are only needed in rather different reasoning scenarios to be included in the generated module. In our approach, we adopted the axiom-based module extraction approach, such that only a small subset of axioms that might be directly related to a reasoning task is included in the generated module. Thus, compactness of our approach is significantly better than the signature-based approach.

The main idea of our approach is that an axiom may specify the "boundaries" for the interpretations of an expression of a symbol. Given the validity test task of an axiom against an ontology, the set of axioms that might be relevant to the reasoning task includes those axioms that may influence the boundaries of symbols used in the testing axiom. Thus, instead of considering the entire ontology to do the validity test, we only need to use the relevant subset (module). Exactness of the method is ensured by discovering all axioms in the original ontology that may directly or indirectly influence boundaries of symbols used in the reasoning task.

The strength of the proposed method includes the following:

(1) Boundary-based module is compact. The experimental results show that modules generated by our method are close to optimal axiom-based modules, and are smaller than modules generated by the signature-based approach.

(2) Most work of computing axiom-based module using the boundary method can be done offline, thus it is suitable for optimizing reasoning performance.

(3) The method does not heavily rely on a specific language; it can be applied to other logic languages that have a Tarski-style model theoretic semantics with only a few modification.

The present paper focuses on the description logic  $\mathcal{EL}^{++}$  [14] which is a notable subset of OWL DL, it aims at high efficiency for reasoning tasks such as subsumption, classification and satisfiability. The advantage of  $\mathcal{EL}^{++}$  is that it combines tractability of the afore mentioned reasoning problems with expressive power that is sufficient for many important applications of ontologies, especially for life science ontologies.

## 2 Related Work

**Signature-based module.** As briefly outlined in Section 1, signature-based module extraction method [9, 10, 11] uses semantic information to extract modules of ontologies and the modules generated are exact. From the definition of signature-based module [11], we can easily see that, giving an axiom  $\alpha$ , module for all symbols in  $\alpha$  in an ontology  $\mathcal{O}$  is also an axiom-based module for  $\alpha$  (for short,  $\alpha$ -module) in  $\mathcal{O}$ . That means signature-based module may also been used for optimizing ontology reasoning. However, it has been shown that this method is in general not compact, thus the mod-

ules generated by it may contain many unnecessary axioms. As the problem of deciding signature-based module is highly undecidable, an approximation algorithm based on the notion of locality is used to compute a module [11].

The example in Table 1 shows the incompactness of the locality-based method as described in [11]. In the example, we want to calculate an  $\alpha$ -module in  $\mathcal{O}_E$ . According to the relationship between axiom-based module and signature-based module, locality-based Mother-module in  $\mathcal{O}_E$  is also an  $\alpha$ -module. Using the signature-based module extraction algorithm<sup>3</sup>, the Mother-module in  $\mathcal{O}_E$  is  $\{(1), (2), (3), (4), (5), (6), (7)\}$ . It is obvious that a minimal  $\alpha$ -module is  $\{(5), (7)\}$ . From this simple example, we can see that although the signature-module can ensure the exactness of reasoning, it is not necessary compact. On the other hand, we will show later the boundary-based approach proposed in this paper can generate compact axiom-based module for this example.

**Table 1.** Fragment of generation ontology

Ontology $\mathcal{O}_E$ :		
(1): Animal $\sqsubseteq$ Object	(2): Person $\sqsubseteq$ Animal	(3): Female $\sqsubseteq \exists \text{has\_Child}.\text{Animal}$
(4): Female $\sqsubseteq$ Animal	(5): Woman $\sqsubseteq$ Person	(6): Woman $\sqsubseteq$ Female
(7): Mother $\sqsubseteq$ Woman	(8): Grandmother $\sqsubseteq$ Mother	(9): Female $\sqcap$ Male $\sqsubseteq \perp$
Axiom $\alpha$ : Mother $\sqsubseteq$ Person		

In [15], authors introduce a variant of locality-based module extraction method which is based on connected reachability from  $\mathcal{EL}^+$  ontologies. Although the reachability-based module is the minimal locality-based module, it still contains many irrelevant axioms for a reasoning task. In addition, the reachability-based module extraction approach can only be applied to  $\mathcal{EL}^+$  ontologies. On the other hand, our approach is versatile to be applied for other description logics.

**Module extraction method using structural information.** The modularization problem has also been addressed in [6], [7] and [8]. Reference [6] presents a partitioning algorithm in order to facilitate the visualization through the ontology. In [7] and [8], authors present methods to extract modules of ontologies for a given signature. All these methods only use structural information of ontologies without providing a characterization for the logical properties of the extracted modules, nor do they establish a notion of exactness of modularization.

**Relevant axiom selection for reasoning.** In [16], authors report a method which can select information relevant to a concept expression in order to enhance the performance of reasoning. The method assumes that the ontology is separated into a set of unfoldable axioms and a set of GCIs. Every concept name and role name appearing in the target concept expression E is relevant to E. The module extraction process is then repeated for unfoldable axioms. Experiments show that the removal of irrelevant information results is a great improvement for the reasoner. However, module extracted using this method still contain many irrelevant axioms for reasoning tasks thus is not compact.

<sup>3</sup> Here we use the semantic locality testing method described in [11], there are also other testing methods recently introduced in [10].

**Justification.** Another related work is justification which is intended to be a debugging technique [17]. Justifications of an entailment are all the minimal set of axioms sufficient to produce an entailment. Let  $\text{JUST}(\alpha, \mathcal{O})$  be the justification for the reasoning task  $\mathcal{O} \models \alpha$ , it is obviously a minimal  $\alpha$ -module in ontology  $\mathcal{O}$ , but it can only be calculated *after* the process of inference, so it cannot be used to optimize reasoning. On the other hand, our approach can produce a close approximation of the minimal module *before* the inference being preformed.

### 3 Preliminaries

A description logic language contains a set of atomic concepts  $(A, B, \dots)$  representing sets of elements, a set of atomic roles  $(r, s, \dots)$  representing binary relations between elements, and a set of individuals  $(a, b, c, \dots)$  representing elements; it also provides constructors for defining the set of (general) roles  $(R, S, \dots)$ , the set of (general) concepts  $(C, D, \dots)$ , and the set of axioms  $(\alpha, \beta, \dots)$  which is an union of role axioms (RBox), terminological axioms (TBox) and assertions (ABox). The semantics for a description logic language is defined using the *Tarski-style* model-theoretic semantics. An interpretation  $\mathcal{I}$  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 that assigns: to every atomic concept  $A$  a set  $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ , to every atomic role  $r$  a set  $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  and every individual  $a$  a set  $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ . The interpretation  $\cdot^{\mathcal{I}}$  is extended to complex roles and concepts via DL-constructors. An interpretation  $\mathcal{I}$  is a model of an ontology  $\mathcal{O}$  if  $\mathcal{I}$  satisfies all axioms in  $\mathcal{O}$ . An ontology  $\mathcal{O}$  implies an axiom  $\alpha$  (written  $\mathcal{O} \models \alpha$ ) if  $\mathcal{I} \models \alpha$  for every model  $\mathcal{I}$  of  $\mathcal{O}$ . Reference [18] provides the detailed syntax and semantics of DL.

In this paper we focus on the DL  $\mathcal{EL}^{++}$ [14], which provides the following main features: top concept  $\top$ , bottom concept  $\perp$ , nominal  $\{a\}$ , conjunction  $C \sqcap D$ , existential restriction  $\exists r.C$ , domain restriction  $\text{dom}(r) \sqsubseteq C$ , range restriction  $\text{ran}(r) \sqsubseteq C$ , general concept inclusion (GCIs)  $C \sqsubseteq D$ , role inclusions (RIs)  $r_1 \circ \dots \circ r_k \sqsubseteq r$ , concept assertion  $C(a)$  and role assertion  $r(a, b)$ .

The signature  $\text{Sig}(\alpha)$  of an axiom  $\alpha$  is the union of role, concept and individual symbols that occurring in  $\alpha$ . The  $\text{Sig}(\mathcal{O})$  of an ontology  $\mathcal{O}$  is defined analogously. Expression  $\text{model}(\mathcal{O})$  represents all models of the ontology  $\mathcal{O}$ .  $\text{model}(\alpha)$  is defined analogously for an axiom  $\alpha$ .

## 4 Axiom-based Module Extraction by Using Boundary

### 4.1 Axiom-based Module

Firstly, we introduce some important definitions and properties of axiom-based module. According to the exactness and compactness requirement, the notion and properties of axiom-based module and minimal axiom-based module are formalized as follows<sup>4</sup>:

<sup>4</sup> If there is no special explanation, the ontology in question is consistent, and the axiom in question is not a tautology or a contradiction axiom (i.e, there is no model for the axiom) in this paper.

**Definition 1 (Axiom-based Module [4])** Let  $\mathcal{O}$  be an ontology and  $\mathcal{O}_1 \subseteq \mathcal{O}$  a subset of axioms in  $\mathcal{O}$ . We say  $\mathcal{O}_1$  is a module for an axiom  $\alpha$  in ontology  $\mathcal{O}$  w.r.t. language  $\mathcal{L}$  (or for short, an  $\alpha$ -module in  $\mathcal{O}$  w.r.t.  $\mathcal{L}$ ) if  $\mathcal{O}_1 \models \alpha$  iff  $\mathcal{O} \models \alpha$ . We denote by  $module(\alpha, \mathcal{O})$  the collection of all  $\alpha$ -module in  $\mathcal{O}$ .

**Definition 2 (Minimal Axiom-based Module)** Let  $\mathcal{O}_1$  be an  $\alpha$ -module in  $\mathcal{O}$  w.r.t.  $\mathcal{L}$ . We say that  $\mathcal{O}_1$  is a minimal module for an axiom  $\alpha$  in ontology  $\mathcal{O}$  w.r.t. language  $\mathcal{L}$  (or for short, a minimal  $\alpha$ -module in  $\mathcal{O}$  w.r.t.  $\mathcal{L}$ ) iff there is no  $\alpha$ -module  $\mathcal{O}_2$  in  $\mathcal{O}$  w.r.t.  $\mathcal{L}$  such that  $\mathcal{O}_2 \subset \mathcal{O}_1$ . We denote by  $minModule(\alpha, \mathcal{O})$  the collection of all minimal  $\alpha$ -modules in  $\mathcal{O}$ .

The minimal  $\alpha$ -module in  $\mathcal{O}$  is sufficient to test the validity of  $\alpha$ . It is  $\emptyset$  when  $\mathcal{O} \not\models \alpha$ , and is not empty otherwise. The unification of all minimal  $\alpha$ -modules in  $\mathcal{O}$  is the set of axioms which provide the necessary information in the reasoning process, which we name as *essential axiom-based module*.

**Definition 3 (Essential Axiom-based Module)** Let  $\mathcal{O}$  be an ontology and  $\alpha$  be an axiom, the essential axiom-based module for  $\alpha$  in  $\mathcal{O}$  (or for short, the essential  $\alpha$ -module in  $\mathcal{O}$ )  $\mathcal{O}_e$  is the unification of all the minimal  $\alpha$ -modules in ontology  $\mathcal{O}$ ,  $\mathcal{O}_e = \bigcup_{\mathcal{O}_m \in minModule(\alpha, \mathcal{O})} \mathcal{O}_m$ .

**Proposition 1 (Properties of Essential Axiom-based Module)** Let  $\mathcal{O}$  be an ontology,  $\alpha$  be an axiom, and  $\mathcal{O}_e$  be the essential  $\alpha$ -module in  $\mathcal{O}$ , then the intersection of any  $\alpha$ -module in  $\mathcal{O}$  and any ontology which includes  $\mathcal{O}_e$  is still an  $\alpha$ -module in  $\mathcal{O}$ .  $\forall \mathcal{O}_1, \mathcal{O}'_e. \mathcal{O}_1 \in module(\alpha, \mathcal{O}) \wedge \mathcal{O}'_e \supseteq \mathcal{O}_e \rightarrow (\mathcal{O}'_e \cap \mathcal{O}_1) \in module(\alpha, \mathcal{O})$ .

## 4.2 Boundary

In this section, some boundary related definitions and propositions will be given. We will firstly formalize the possible interpretation and boundary of symbols, and then show that every symbol has boundary in DL  $\mathcal{EL}^{++}$ .

**Definition 4 (Possible Interpretation of Symbols)** Let  $\mathcal{O}$  be an ontology,  $\alpha$  be an axiom, and  $s$  be a symbol in  $\mathcal{O}$ , i.e.,  $s \in Sig(\mathcal{O})$ . Let  $\mathfrak{M}$  be the set of all models of  $\mathcal{O}$ , a possible interpretation of  $s$  is a function  $f_s^\mathcal{O} : \mathfrak{M} \rightarrow S$  where  $S$  is a set, such that for any  $\mathcal{I} \in \mathfrak{M}$ , let  $\mathcal{I}'$  be a new interpretation obtained by letting  $s^{\mathcal{I}'} = X$  and keep interpretations of other symbols unchanged, denoted by  $\mathcal{I} \xrightarrow{s^{\mathcal{I}'}=X} \mathcal{I}'$ , then  $\mathcal{I}'$  is a model of  $\mathcal{O}$  iff  $X \in f_s^\mathcal{O}(\mathcal{I})$ . We call  $f_s^\mathcal{O}(\mathcal{I})$  the possible interpretation of  $s$  in  $\mathcal{O}$ .

Let  $\mathfrak{M}^{+Y}$  be a set of model of  $\mathcal{O} \cup \{\alpha\}$  obtained by setting interpretation of symbols in  $Sig(\alpha) \setminus (Sig(\mathcal{O}) \cap Sig(\alpha))$  to a value  $Y = (Y_1, \dots, Y_i, \dots, Y_n)$  and set other symbols interpretation to all the available interpretations. If there exists a  $\mathfrak{M}^{+Y}$  which satisfies: for all model  $\mathcal{I}^{+Y} \in \mathfrak{M}^{+Y}$ , there exists a model  $\mathcal{I} \in \mathfrak{M}$ ,  $f_s^\mathcal{O}(\mathcal{I}) \neq f_s^{\mathcal{O} \cup \{\alpha\}}(\mathcal{I}^{+Y})$ , we say  $\alpha$  influences the possible interpretation of  $s$  through  $\mathcal{O}$ , in no misunderstanding situation,  $\mathcal{O}$  can be omitted.

Axiom set influences the possible interpretation of  $s$  in  $\mathcal{O}$  contains all axioms  $\alpha$  which influences possible interpretation of  $s$  through any subset of  $\mathcal{O} \setminus \{\alpha\}$ , denoted by  $IS(s, \mathcal{O})$ .

Obviously, every symbol in  $\mathcal{EL}^{++}$  ontologies has a possible interpretation.

**Proposition 2 (Existence of Possible Interpretation)** *Let  $\mathcal{O}$  be an ontology,  $s$  be a symbol in  $\mathcal{O}$ , then there must exist the possible interpretation of  $s$  in  $\mathcal{O}$ .*

*Proof.* In description logic  $\mathcal{EL}^{++}$ , for any symbol  $s$ , there must exist a function  $f(\mathcal{I}) = f(s_1^{\mathcal{I}}, \dots, s_i^{\mathcal{I}}, \dots, s_k^{\mathcal{I}})$ ,  $s_i \in \text{Sig}(\mathcal{O}) \setminus s, \forall \mathcal{I} \in \text{Model}(\mathcal{O})$  which satisfies the definition of the possible interpretation. Thus there must exist possible interpretation of symbol  $s$ .

The possible interpretation has the following properties:

**Proposition 3 (Properties of Possible Interpretation)** *Let  $\mathcal{O}$  be an ontology,  $\alpha$  be an axiom,  $s$  be a symbol in  $\alpha$ , then  $\mathcal{O} \models \alpha$  if  $\forall \mathcal{I} \in \text{model}(\mathcal{O}), f_s^{\mathcal{O}}(\mathcal{I}) \subseteq f_s^{\alpha}(\mathcal{I})$ .*

*Proof.*  $\mathcal{O} \models \alpha$  iff  $\text{model}(\mathcal{O}) \subseteq \text{model}(\alpha)$ , according to the definition and existence proposition of the possible interpretation,  $\forall \mathcal{I} \in \text{model}(\mathcal{O})$ , if  $f_s^{\mathcal{O}}(\mathcal{I}) \subseteq f_s^{\alpha}(\mathcal{I})$ , then  $\mathcal{I}$  is also a model of  $\alpha$ , so  $\text{model}(\mathcal{O}) \subseteq \text{model}(\alpha)$ ,  $\mathcal{O} \models \alpha$ .

**Definition 5 (Boundary of Symbols)** *Let  $\mathcal{O}$  be an ontology and  $s$  be a symbol<sup>5</sup> in  $\mathcal{O}$ , i.e.,  $s \in \text{Sig}(\mathcal{O})$ . Let  $\mathfrak{M}$  be the set of all models of  $\mathcal{O}$ . Boundary of  $s$  is a pair  $(l_s^{\mathcal{O}}, u_s^{\mathcal{O}})$ ,  $l_s^{\mathcal{O}}$  and  $u_s^{\mathcal{O}}$  are two functions:  $\mathfrak{M} \rightarrow (S_l, S_u)$  where  $S_l$  and  $S_u$  are two sets, such that (i) for any  $\mathcal{I} \in \mathfrak{M}$ ,  $S_l = l_s^{\mathcal{O}}(\mathcal{I}) \subseteq \Delta^{\mathcal{I}}$  or  $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ ,  $S_u = u_s^{\mathcal{O}}(\mathcal{I}) \subseteq \Delta^{\mathcal{I}}$  or  $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}6}$ ; (ii) For a given  $\mathcal{I}$  and  $X \subseteq \Delta^{\mathcal{I}}$ , let  $\mathcal{I}'$  be a new interpretation obtained from  $\mathcal{I}$  by letting  $s^{\mathcal{I}'} = X$  and keep other symbols interpretations unchanged, denoted by  $\mathcal{I} \xrightarrow{s^{\mathcal{I}'}=X} \mathcal{I}'$ , then  $\mathcal{I}'$  is a model of  $\mathcal{O}$  iff  $l_s^{\mathcal{O}}(\mathcal{I}) \subseteq X \subseteq u_s^{\mathcal{O}}(\mathcal{I})$ .*

We call  $l_s^{\mathcal{O}}$  and  $u_s^{\mathcal{O}}$  the lower and upper boundary of  $s$  in  $\mathcal{O}$ <sup>7</sup>, when a lower boundary is the empty set and an upper boundary is the universal set, they are called trivial<sup>8</sup>. Otherwise, it is nontrivial.

From the definition of possible interpretation and boundary, we can see that the possible interpretation and the boundary for a symbol  $s$  in an ontology  $\mathcal{O}$  are mutual determined,  $f_s^{\mathcal{O}} \Leftrightarrow (l_s^{\mathcal{O}}, u_s^{\mathcal{O}})$ .

**Proposition 4 (Existence of Boundary of symbols)** *Let  $\mathcal{O}$  be an  $\mathcal{EL}^{++}$  ontology and  $s$  be a symbol in  $\mathcal{O}$ ,  $s \in \text{Sig}(\mathcal{O})$ , then there must exist lower boundary  $l_s^{\mathcal{O}}$  and upper boundary  $u_s^{\mathcal{O}}$ .*

*Proof.* Symbol  $s$  has the possible interpretation  $f_s^{\mathcal{O}}(\mathcal{I})$  in  $\mathcal{O}$ , the lower and upper boundary of  $s$  can be constructed using the following method: If there exists  $f_1(\mathcal{I})$  and  $f_2(\mathcal{I})$  which satisfies:  $X \in f_s^{\mathcal{O}}(\mathcal{I})$  iff  $f_1(\mathcal{I}) \subseteq X \subseteq f_2(\mathcal{I})$ ,  $\forall \mathcal{I} \in \text{model}(\mathcal{O})$ , then  $l_s^{\mathcal{O}} = f_1(\mathcal{I})$ ,  $u_s^{\mathcal{O}} = f_2(\mathcal{I})$ ; else  $l_s^{\mathcal{O}} = f_s^{\mathcal{O}}(\mathcal{I})$ ,  $u_s^{\mathcal{O}} = f_s^{\mathcal{O}}(\mathcal{I})$ .

<sup>5</sup> By default, symbol  $s$  is not  $\top$  or  $\perp$  in this paper.

<sup>6</sup> If  $s$  is a concept symbol,  $S_l \subseteq \Delta^{\mathcal{I}}$ ,  $S_u \subseteq \Delta^{\mathcal{I}}$ ; if  $s$  is a role symbol,  $S_l \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ ,  $S_u \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$

<sup>7</sup> In the no misunderstanding situation,  $\mathcal{O}$  can be omitted from  $l_s^{\mathcal{O}}$  and  $u_s^{\mathcal{O}}$

<sup>8</sup> In description logics, if  $s$  is a concept symbol, the trivial lower and upper boundary of a symbol are  $\emptyset$  and  $\Delta^{\mathcal{I}}$ , if  $s$  is a role symbol, the trivial lower and upper boundary of a symbol are  $\emptyset$  and  $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ .

It is an extreme case that the lower boundary and upper boundary are equal, in  $\mathcal{EL}^{++}$  ontologies, the lower boundary and upper boundary are usually unequal. For instance, let  $\mathcal{O}$  be  $\{C \sqsubseteq D\}$ ,  $s$  be  $C$ , then  $l_s^{\mathcal{O}} = \emptyset, u_s^{\mathcal{O}} = D^{\mathcal{I}}, \forall \mathcal{I} \in \text{model}(\mathcal{O})$ . Note that we do not give a method for computing the boundary functions as our module extraction method only needs the existence of boundaries, another reason is that calculation of boundaries is very complex.

When  $\mathcal{O}$  is empty, the lower and upper boundary of  $s$  is trivial. When  $\mathcal{O}$  is not empty, the boundary of  $s$  are determined by some axioms in the ontology. For example, in ontology  $\mathcal{O}_E$ , we extract two axioms:  $\alpha : \text{Grandmother} \sqsubseteq \text{Mother} \quad \beta : \text{Mother} \sqsubseteq \text{Woman}$ . For symbol Grandmother, axiom  $\alpha$  means the interpretation of Grandmother is included in the interpretation of Mother,  $u_{\text{Grandmother}}$  is changed from  $\Delta^{\mathcal{I}}$  to  $\text{Mother}^{\mathcal{I}}$  and the lower boundary is not changed, we say  $\alpha$  is relevant to the upper boundary of Grandmother. On the another hand, for symbol Mother,  $\alpha$  means the interpretation of Mother must include the interpretation of Grandmother, the  $l_{\text{Mother}}$  is changed from  $\emptyset$  to  $\text{Grandmother}^{\mathcal{I}}$  and the upper boundary is not changed, we say axiom  $\alpha$  is relevant to the lower boundary of Mother. We can also notice that  $\beta$  is upper boundary relevant to symbol Grandmother through  $\alpha$ , as  $\beta$  influences the  $u_{\text{Grandmother}}$  by influencing the possible interpretation of Mother. We call axioms like  $\alpha$  direct boundary relevant axioms and axioms like  $\beta$  indirect boundary relevant axioms for symbol Grandmother. The boundary relevant axiom is formalized as follows:

**Definition 6 (Boundary Relevant Axiom)** *Let  $\mathcal{O}$  be an ontology,  $s$  be a symbol in  $\mathcal{O}$ ,  $\alpha$  be an axiom which cannot be implied from  $\mathcal{O}$ ,  $\mathcal{O} \not\models \alpha$ ,  $\mathcal{O}^*$  be an ontology which add  $\alpha$  into  $\mathcal{O}$ ,  $\mathcal{O}^* = \mathcal{O} \cup \{\alpha\}$ . Let  $\mathfrak{M}^{+Y}$  be a set of model of  $\mathcal{O}^*$  obtained by setting interpretation of symbols in  $\text{Sig}(\alpha) \setminus (\text{Sig}(\mathcal{O}) \cap \text{Sig}(\alpha))$  to a value  $Y = (Y_1, \dots, Y_i, \dots, Y_n)$  and set other symbols interpretation to all the available interpretations. If there exists a  $\mathfrak{M}^{+Y}$  which satisfies: for all model  $\mathcal{I}^{+Y} \in \mathfrak{M}^{+Y}$ , there exists a model  $\mathcal{I} \in \mathfrak{M}$ ,  $l(u)_s^{\mathcal{O}}(\mathcal{I}) \neq l(u)_s^{\mathcal{O}^*}(\mathcal{I}^{+Y})$ , we say axiom  $\alpha$  is relevant to the lower(upper) boundary of  $s$  through  $\mathcal{O}$ , denoted by  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \mathcal{O}, \alpha) = 1$ ; else we say axiom  $\alpha$  is irrelevant to the lower(upper) boundary of  $s$  through  $\mathcal{O}$ , denoted by  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \mathcal{O}, \alpha) = 0$ . Furthermore, relevant axioms can be divided into direct and indirect boundary relevant axioms:*

(1) **Direct boundary relevant axioms.** *If  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \emptyset, \alpha) = 1$ , then we say axiom  $\alpha$  is directly relevant to the lower(upper) boundary of  $s$ ,  $\mathcal{DL}(\mathcal{U})\mathcal{B}(s, \mathcal{O})$  is used to denote all these direct relevantly axioms in  $\mathcal{O}$ ;*

(2) **Indirect boundary relevant axioms.** *If  $\alpha$  is not direct boundary relevant to the lower(upper) boundary of  $s$ ,  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \emptyset, \alpha) = 0$ , and  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \mathcal{O}, \alpha) = 1$ , then we say  $\alpha$  is indirectly relevant to the lower boundary of  $s$  through  $\mathcal{O}$ .  $\mathcal{IL}(\mathcal{U})\mathcal{B}(s, \mathcal{O})$  is used to denote all these indirect relevant axioms in  $\mathcal{O}$ .*

**Proposition 5 (Properties of Boundary Relevance)** *Let  $\mathcal{O}$  be an ontology,  $s$  be a symbol in  $\mathcal{O}$ ,  $s \in \text{Sig}(\mathcal{O})$ ,  $\alpha$  be an axiom which cannot be implied from  $\mathcal{O}$ ,  $\mathcal{O} \not\models \alpha$ ,  $\mathcal{O}^+$  be an ontology which includes  $\mathcal{O}$ ,  $\mathcal{O} \subseteq \mathcal{O}^+$ ,  $\mathcal{O}^-$  be an ontology which is included by  $\mathcal{O}$ ,  $\mathcal{O}^- \subseteq \mathcal{O}$ . (1) If  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \mathcal{O}, \alpha) = 1$ , then  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \mathcal{O}^+, \alpha) = 1$ ; (2) If  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \mathcal{O}, \alpha) = 0$ , then  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \mathcal{O}^-, \alpha) = 0$ .*

*Proof.* We prove the lower boundary relevance, and the upper boundary relevance can be proved analogously. (1) Since  $\mathcal{L}\mathcal{BR}(s, \mathcal{O}, \alpha) = 1$  and  $\mathcal{O} \subseteq \mathcal{O}^+$ ,  $\mathcal{O}^+$  may contain ad-

ditional axioms which influence the lower boundary for symbol  $s$ , thus  $\mathcal{LBR}(s, \mathcal{O}^+, \alpha) = 1$ . (2) Since  $\mathcal{LBR}(s, \mathcal{O}, \alpha) = 0$  and  $\mathcal{O}^- \subseteq \mathcal{O}$ ,  $\mathcal{O}^-$  does not contain additional axioms which influence the lower boundary for symbol  $s$ , thus  $\mathcal{LBR}(s, \mathcal{O}^-, \alpha) = 0$ .

For a symbol in ontology, we can partition the ontology into three parts: upper boundary relevant, lower boundary relevant and irrelevant axiom set.

**Definition 7 (Boundary Relevant Axiom Set)** Let  $\mathcal{O}$  be an ontology,  $s$  be a symbol in  $\mathcal{O}$ ,  $s \in \text{Sig}(\mathcal{O})$ .

(1) **Lower(Upper) boundary relevant.** Axiom set relevant to lower(upper) boundary of  $s$  in  $\mathcal{O}$  contains all axioms  $\alpha$  which are relevant to lower(upper) boundary of  $s$  through any subset of  $\mathcal{O} \setminus \{\alpha\}$ , denoted by  $\mathcal{L}(\mathcal{U})\mathcal{R}(s, \mathcal{O})$ . According to proposition 5,  $\mathcal{L}(\mathcal{U})\mathcal{R}(s, \mathcal{O}) = \{\alpha \mid \mathcal{L}(\mathcal{U})\mathcal{BR}(s, \mathcal{O} \setminus \{\alpha\}, \alpha) = 1\}$ ;

(2) **Boundary irrelevant.** Axiom set irrelevant to the boundary of  $s$  in  $\mathcal{O}$  contains all axioms  $\alpha$  which are irrelevant to boundary of  $s$  through any subset of  $\mathcal{O} \setminus \{\alpha\}$ , denoted by  $\mathcal{IR}(s, \mathcal{O})$ .  $\mathcal{IR}(s, \mathcal{O}) = \mathcal{O} \setminus (\mathcal{LR}(s, \mathcal{O}) \cup \mathcal{UR}(s, \mathcal{O}))$ .

Based on the definition 7,  $\mathcal{LR}(s, \mathcal{O})$  and  $\mathcal{UR}(s, \mathcal{O})$  contain all axioms which change the lower and upper boundary of  $s$  in  $\mathcal{O}$ , axioms in  $\mathcal{IR}(s, \mathcal{O})$  do not influence the boundary of  $s$ . Boundary relevant axiom set has following properties.

**Proposition 6 (Properties of Boundary Relevant Axiom Set)** Let  $\mathcal{O}$  be an ontology,  $s$  be a symbol in  $\mathcal{O}$ ,  $s \in \text{Sig}(\mathcal{O})$ . (1)  $\mathcal{L}(\mathcal{U})\mathcal{R}(s, \mathcal{O}) = \mathcal{DL}(\mathcal{U})\mathcal{R}(s, \mathcal{O}) \cup \mathcal{IL}(\mathcal{U})\mathcal{R}(s, \mathcal{O})$ ; (2)  $\mathcal{LR}(s, \mathcal{O}) \cup \mathcal{UR}(s, \mathcal{O}) = \mathcal{IS}(s, \mathcal{O})$ .

### 4.3 Boundary-based Module

In this section, we will firstly give the definition of useful and useless symbols, then formalize the relationship between boundary relevant axiom set and axiom-based module.

**Definition 8 (Useful and Useless Symbols)** Let  $\mathcal{O}$  be an ontology,  $\alpha$  be an axiom implied by  $\mathcal{O}$ ,  $\mathcal{O} \models \alpha$ ,  $s$  be a symbol in  $\alpha$ ,  $\mathcal{I}$  be an model of  $\mathcal{O}$ . let  $\mathcal{I}'$  be a new interpretation obtained from  $\mathcal{I}$  by letting  $s^{\mathcal{I}'} = X$  and keep other symbols interpretations unchanged, denoted by  $\mathcal{I} \xrightarrow{s^{\mathcal{I}'}=X} \mathcal{I}'$ . If for any model  $\mathcal{I}$  of  $\mathcal{O}$ , no matter what the  $X$  is,  $\mathcal{I}'$  is still a model of  $\alpha$ , we say  $s$  is useless symbol in  $\alpha$  w.r.t  $\mathcal{O}$ ; else  $s$  is useful symbol in  $\alpha$  w.r.t  $\mathcal{O}$ .  $\mathcal{US}(\alpha, \mathcal{O})$  is used to denote useful symbols set in  $\alpha$  w.r.t  $\mathcal{O}$ .

Intuitively, symbol  $s$  is useless in  $\alpha$  w.r.t  $\mathcal{O}$ , that means  $s$  has nothing to do with the reasoning process  $\mathcal{O} \models \alpha$ . For instance, in  $\mathcal{O}_E$ , let axiom  $\alpha$  be Female  $\sqcap$  Male  $\sqsubseteq$  Grandmother, then symbol Grandmother is an useless symbol in  $\alpha$  w.r.t  $\mathcal{O}_E$  as Female  $\sqcap$  Male  $\sqsubseteq \perp$ . On the other hand, symbol  $s$  is useful in  $\alpha$  w.r.t  $\mathcal{O}$ , that means  $s$  has something to do with the reasoning process  $\mathcal{O} \models \alpha$ . In the above example, it is easy to see that symbols Female and Male are useful symbols in  $\alpha$  w.r.t  $\mathcal{O}_E$ . Generally speaking, axioms representing reasoning tasks in practice are meaningful, which means they only contain useful symbols. For  $\mathcal{O} \models \alpha$ , in the ontology  $\mathcal{O}$ , axiom set which influences the same boundary of an useful symbol as axiom  $\alpha$  does provides sufficient condition

to compute the reasoning process, in other words, the axiom set is an  $\alpha$ -module in  $\mathcal{O}$ . The relationship between boundary relevant axiom set and axiom-based module can be formalized as follows:

**Proposition 7 (Boundary-based Module)** *Let  $\mathcal{O}$  be an ontology,  $\alpha$  be an axiom and  $s$  be an useful symbol in  $\alpha$ . (1) If  $\mathcal{L}(\mathcal{U})\mathcal{BR}(s, \emptyset, \alpha) = 1$  and  $\mathcal{U}(\mathcal{L})\mathcal{BR}(s, \emptyset, \alpha) = 0$ , then axiom set relevant to lower(upper) boundary of  $s$  in  $\mathcal{O}$  is the superset of essential  $\alpha$ -module in  $\mathcal{O}$ ,  $\mathcal{O}_e \subseteq \mathcal{L}(\mathcal{U})\mathcal{R}(s, \mathcal{O})$ ; (2) If  $\mathcal{L}\mathcal{BR}(s, \emptyset, \alpha) = 1$  and  $\mathcal{U}\mathcal{BR}(s, \emptyset, \alpha) = 1$ , then axiom set relevant to lower or upper boundary of  $s$  in  $\mathcal{O}$  is the superset of essential  $\alpha$ -module in  $\mathcal{O}$ ,  $\mathcal{O}_e \subseteq \mathcal{LR}(s, \mathcal{O}) \cup \mathcal{UR}(s, \mathcal{O})$ .*

*Proof.*  $\mathcal{O} \models \alpha$  if  $\forall \mathcal{I} \in \text{model}(\mathcal{O}), f_s^{\mathcal{O}}(\mathcal{I}) \subseteq f_s^{\alpha}(\mathcal{I})$ . And according to the definition and existence proposition of the boundary,  $s$  has the lower and upper boundary, and the boundary and the possible interpretation are mutual determined, so boundary of  $s$  in  $\mathcal{O}$  and  $\alpha$  should satisfy:  $\forall \mathcal{I} \in \text{model}(\mathcal{O}), l_s^{\alpha} \subseteq l_s^{\mathcal{O}}, u_s^{\mathcal{O}} \subseteq u_s^{\alpha}$ . When  $\alpha$  is only relevant to the lower boundary of  $s$ ,  $u_s^{\mathcal{O}} \subseteq u_s^{\alpha}$  is always satisfied as  $u_s^{\alpha} = \Delta^{\mathcal{I}}$  or  $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ . It is easy to observe all axioms which influence the lower boundary of  $s$  in  $\mathcal{O}$  provide the necessary information to test if  $\forall \mathcal{I} \in \text{model}(\mathcal{O}), l_s^{\alpha} \subseteq l_s^{\mathcal{O}}$ . Hence, the lower boundary relevant axiom set includes  $\mathcal{O}_e$ . Other situation can be proved analogously.

We use  $\text{boundaryModule}(s, \alpha, \mathcal{O})$  to denote the  $s$  boundary-based  $\alpha$ -module in  $\mathcal{O}$ . Based on the proposition 7, we give a method to calculate the minimal boundary-based module.

**Proposition 8 (Minimal Boundary-based Module)** *Let  $\mathcal{O}$  be an ontology and  $\alpha$  be an axiom, then intersection of all axiom sets relevant to the same boundary of symbols of  $\mathcal{US}(\alpha, \mathcal{O})$  as  $\alpha$  does is an  $\alpha$ -module in  $\mathcal{O}$ ,  $\bigcap_{s \in \mathcal{US}(\alpha, \mathcal{O})} \text{boundaryModule}(s, \alpha, \mathcal{O}) \in \text{module}(\alpha, \mathcal{O})$ , we name it the minimal boundary-based  $\alpha$ -module in  $\mathcal{O}$ , denoted by  $b\text{Mod}(\alpha, \mathcal{O})$ .*

The above proposition 8 shows that a reasoning task contains more useful symbols, the boundary-based minimal module will not be larger. On the contrary, the volume of signature-based module becomes bigger when the reasoning task contains more symbols. This means the boundary-based minimal module is suitable for optimizing reasoning, especially for complex tasks which contain many symbols.

#### 4.4 Calculation of the Boundary Relevant Axiom Set

Definition of boundary relevant axiom set can be used to compute the axiom set relevant to the boundary of  $s$  in  $\mathcal{O}$ . Using this method, we have to compute the boundary. As it is very complex, direct and indirect boundary relevance are used to calculate the boundary relevant axiom set.

As  $\mathcal{L}(\mathcal{U})\mathcal{B}(s, \mathcal{O}) = \mathcal{DL}(\mathcal{U})\mathcal{B}(s, \mathcal{O}) \cup \mathcal{IL}(\mathcal{U})\mathcal{B}(s, \mathcal{O})$ , calculation of the axiom set relevant to the boundary of  $s$  in  $\mathcal{O}$  can be divided into computation of direct and indirect boundary relevant axiom set. We firstly introduce properties of indirect relevance, then describe the algorithm for computing boundary relevant axiom set.

**Proposition 9 (Properties of Indirect Relevance)** *Let  $\mathcal{O}$  be an ontology,  $s$  be a symbol in  $\mathcal{O}$ ,  $\beta$  be an axiom which is indirectly relevant to the lower(upper) boundary of  $s$ . Then there must exist an axiom  $\alpha$  in  $\mathcal{O}$ ,  $\alpha$  is directly relevant to the lower(upper) boundary of  $s$ , and  $\beta$  changes the lower(upper) boundary of  $s$  by influencing the possible interpretation of some symbol in  $Sig(\alpha) \setminus s$ .*

To calculate the indirect boundary relevant axiom set of  $s$  in  $\mathcal{O}$ , we can analyze every axiom  $\beta$  which influences the possible interpretation of symbols in axioms which are directly relevant to the boundary of  $s$ , i.e.,  $\beta \in \cup_{\alpha_i \in \mathcal{DLB}(s, \mathcal{O})} \mathcal{IS}(Sig(\alpha_i) \setminus s, \mathcal{O})$ . We use algorithm 1 to compute all axioms relevant to the lower boundary of  $s$  in  $\mathcal{O}$ , and axioms relevant to the upper boundary can be computed analogously.

---

**Algorithm 1** Computing lower boundary relevant axiom set

---

**Input:**  $s, \mathcal{O}$     **Output:**  $\mathcal{L}(s, \mathcal{O})$

1. In  $\mathcal{O}$ , calculating axioms which are directly relevant to the lower boundary of  $s$ , put them into  $S$  and  $\mathcal{L}(s, \mathcal{O})$ ; (**Testing the direct boundary relevance between  $s$  and all axioms in  $\mathcal{O}$** )
  2. For each axiom  $\alpha$  in  $S$ , calculating axioms which are indirectly relevant to the lower boundary of  $s$  by influencing the possible interpretation of any symbols in  $Sig(\alpha) \setminus s$ , put them into  $\mathcal{L}(s, \mathcal{O})$ ; (**Transformed into direct boundary relevance testing**)
- 

**Proposition 10 (Exactness of Algorithm 1)** *For every ontology  $\mathcal{O}$  and symbol  $s$  in  $\mathcal{O}$ . The algorithm 1 computes a correct  $\mathcal{LR}(s, \mathcal{O})$ .*

*Proof.* We have to show that (1) algorithm 1 terminates for every input  $\mathcal{O}$  and  $s$ , and (2)  $\mathcal{L}(s, \mathcal{O})$  is  $\mathcal{LR}(s, \mathcal{O})$ . (1) Termination of the algorithm follows from the fact that there is a certain number of axioms which is directly relevant to the lower boundary of  $s$  in  $\mathcal{O}$ . Thus the algorithm can be terminated. (2) Any axiom which is indirectly relevant to the lower boundary of  $s$  must influence the possible interpretation of some symbols in axioms which are directly relevant to the lower boundary of  $s$ . Hence,  $\mathcal{L}(s, \mathcal{O})$  can be  $\mathcal{LR}(s, \mathcal{O})$  by using a proper analysis method.

When using algorithm 1, we firstly give a method for testing the direct boundary relevance, then use an elimination method to testing which axiom is indirectly relevant to the boundary of  $s$  by influencing possible interpretation of symbols in  $\alpha$ . As we know, axioms which are indirectly relevant to the boundary of  $s$  by influencing possible interpretation of symbols in  $\alpha$  are contained in the boundary relevant axiom set of symbols in  $Sig(\alpha) \setminus s$ , we can analyze these axioms and remove those irrelevant to the boundary of  $s$ , the remains are considered as indirectly relevant to the boundary of  $s$  by influencing possible interpretation of symbols in  $\alpha$ , the analysis process can be transformed into the problem of testing direct boundary relevance.

**Proposition 11 (Testing Direct Boundary Relevance)** *Let  $\alpha$  be an axiom and  $s$  be a symbol in  $\alpha$ ,  $s \in Sig(\alpha)$ . We set interpretation of  $s$  to the empty set (universe set),  $\alpha$  is changed to  $\alpha^*$ , and test whether  $\alpha$  implies  $\alpha^*$ . (1) If  $\alpha \models \alpha^*$ , then  $\alpha$  is not directly relevant to the lower(upper) boundary of  $s$ ; (2) If  $\alpha \not\models \alpha^*$ , then  $\alpha$  is directly relevant to the lower(upper) boundary of  $s$ .*

*Proof.* We prove the correctness of testing the direct lower boundary relevance, and the upper situation can be proved using the similar way. (1) If  $\alpha \models \alpha^*$ , then  $model(\alpha) \subseteq model(\alpha^*)$ , which means, for any model  $\mathcal{I}$  of  $\alpha$ , when all interpretation of other symbols in  $Sig(\alpha) \setminus s$  remain unchanged,  $s^{\mathcal{I}}$  can be the empty set. Thus the lower boundary of  $s$  is not changed, and  $\alpha$  is not directly relevant to the lower boundary of  $s$ . (2) If  $\alpha \not\models \alpha^*$ , this means, there exists at least one model  $\mathcal{I}^*$  of  $\alpha$ , when interpretation of other symbols in  $Sig(\alpha) \setminus s$  is given a special value,  $s^{\mathcal{I}^*}$  cannot be the empty set. Hence  $\alpha$  is directly relevant to the lower boundary of  $s$ .

**Proposition 12 (Indirect Relevance Testing  $\Rightarrow$  Direct Relevance Testing)** *Let  $\alpha$  be an axiom,  $s, t$  be two symbols in  $\alpha$ ,  $s, t \in Sig(\alpha)$ , and  $\alpha$  is directly relevant to the boundary of  $s$ . If  $\alpha$  is not directly relevant to the lower(upper) boundary of  $t$ , then axioms which are not relevant to the lower(upper) boundary of  $t$  are not indirectly relevant to the boundary of  $s$  by influencing the possible interpretation of  $t$  in  $\alpha$ .*

*Proof.* We prove the correctness of testing the indirect lower boundary relevance, and the upper situation can be proved using the similar way. Axiom  $\alpha$  is not directly relevant to the lower boundary of  $t$ , so for any model  $\mathcal{I}$  of  $\alpha$ , when change  $t^{\mathcal{I}}$  to the empty set, interpretation of other symbols can keep unchanged, and  $s^{\mathcal{I}}$  remain the same. And if an axiom is not relevant to the lower boundary of  $t$ , then  $t^{\mathcal{I}}$  can be the empty set in the model of this axiom. Thus, the axiom is not indirectly relevant to the boundary of  $s$  by influencing the possible interpretation of  $t$  in  $\alpha$ .

According to the proposition 12, we use the complement of axioms which are not indirectly relevant to get approximate axiom set which is indirect boundary relevant. To calculate the approximate axiom set relevant to the boundary of symbol  $s$  in ontology  $\mathcal{O}$ , axiom set  $S_1$  which is directly relevant to the boundary of symbol  $s$  and axiom set  $S_2$  which is indirectly relevant to the boundary of  $s$  should be computed. Firstly,  $S_1$  can be easily obtained by testing the direct relevance between  $s$  and all axioms in  $\mathcal{O}$ , then calculation for  $S_2$  can be changed to calculation of axiom set which is direct and indirect to the boundary of other symbols, the process is repeated until the boundary relevant axiom set is obtained.

Using boundary-based module extraction method, we can get an axiom-based module far smaller than signature-based module in ontology  $\mathcal{O}_E$ . Let  $\mathcal{I}$  by any model of  $\mathcal{O}_E$ ,  $u_{Mother}$  is changed from  $\Delta^{\mathcal{I}}$  to  $Woman^{\mathcal{I}}$  by  $\{(1) \sim (7), (9)\}$ ;  $l_{Person}$  is changed from  $\emptyset$  to  $Woman^{\mathcal{I}}$  by  $\{(5), (7), (8)\}$ . Cause Mother and Person are useful symbols in  $\alpha$  w.r.t  $\mathcal{O}_E$ , and axiom  $\alpha$  influence  $u_{Mother}$  and  $l_{Person}$ ,  $\alpha$ -module in  $\mathcal{O}_E$  is the intersection of axiom sets which change  $u_{Mother}$  and  $l_{Person}$ ,  $\{(1) \sim (7), (9)\} \cap \{(5), (7), (8)\} = \{(5), (7)\}$ , which is the minimal axiom-based module for  $\alpha$  in  $\mathcal{O}_E$ .

Note: As testing the direct relevance involves computing many implication problem, and checking for implication in description logics is, theoretically, a difficult problem, we can firstly compute all the direct boundary relevant relationship between symbols and axioms offline, then the axiom-based module task can be easily and fast obtained.

## 5 Implementation in $\mathcal{EL}^{++}$ Description Logics Language

We use Pellet<sup>9</sup> and OWLAPI<sup>10</sup> to implement the boundary-based module extraction method in DL  $\mathcal{EL}^{++}$ . As analyzed above, calculation for axiom-based module can be transformed into testing the direct boundary relevance. Proposition 11 is used to testing the direct boundary relevance. Given an axiom  $\alpha$ , and a symbol  $s$  in  $\alpha$ , the testing approach is divided into three situations according to the type of symbol  $s$ .

**Symbol  $s$  is a concept  $C$ .** We set  $C$  to the top concept  $\top$ ,  $\alpha_{C^{\rightarrow\top} \Delta^x}$ , and set  $C$  to the bottom concept  $\perp$ ,  $\alpha_{C^{\rightarrow\perp} \emptyset}$ . Then we test whether  $\alpha$  implies  $\alpha_{C^{\rightarrow\top} \Delta^x}$  and  $\alpha_{C^{\rightarrow\perp} \emptyset}$ ; if  $\alpha \models \alpha_{C^{\rightarrow\top} \Delta^x} (\alpha_{C^{\rightarrow\perp} \emptyset})$ , then  $\alpha$  is not directly relevant to the upper(lower) boundary of  $C$ ; else if  $\alpha \not\models \alpha_{C^{\rightarrow\top} \Delta^x} (\alpha_{C^{\rightarrow\perp} \emptyset})$ , then  $\alpha$  is directly relevant to the upper(lower) boundary of  $C$ .

**Symbol  $s$  is an individual.** In the DL  $\mathcal{EL}^{++}$ , axiom types which contains individual are  $a : C$  and  $r(a, b)$ <sup>11</sup>. In these axioms, upper boundary of the interpretation of individuals are influenced, so when symbol  $s$  is an individual,  $\alpha$  is directly relevant to the upper boundary of  $s$ .

**Symbol  $s$  is a role  $r$ .** As there are not constants denoting the role whose interpretation is the universal or the empty set in the description logics, proposition 11 cannot be used directly. In DL  $\mathcal{EL}^{++}$ , Axiom types which contain role are  $r(a, b)$ ,  $RBox$ , and complex concept  $\exists r.C$ . If  $\alpha$  is  $r(a, b)$ ,  $\alpha$  is directly relevant to the lower boundary of  $r$ ; if  $\alpha$  is  $r_1 \circ \dots \circ r_k \sqsubseteq r$ , then  $\alpha$  is directly relevant to the lower boundary of  $r$ , and directly relevant to the upper boundary of  $r_i, 1 \leq i \leq k$ ; if  $\alpha$  is  $dom(r) \sqsubseteq C$  or  $ran(r) \sqsubseteq C$ , then  $\alpha$  influences the lower and upper boundary of  $r$ ; if  $r$  belongs to complex concept, we change the complex concept which contains role into top or bottom concept according to the semantics of constructors,  $\alpha_{r^{\rightarrow\top} \Delta^x \times \Delta^x}$ ,  $\alpha_{r^{\rightarrow\perp} \emptyset \times \emptyset}$ , then concept boundary testing method is used to testing boundary relevance for  $r$ .

## 6 Experiment

In this section, we will show that the boundary-based module is small enough for optimizing ontology reasoning in DL  $\mathcal{EL}^{++}$ , we compare the minimal boundary-based module ( $bMod$ ) towards the standard axiom-based module: minimal axiom-based module ( $mMod$ ) and essential axiom-based module ( $eMod$ )<sup>12</sup>, in addition, Grau's Locality-based module ( $LoMod$  and  $UpMod$  [9])<sup>13</sup> is also evaluated. A set of well-known  $\mathcal{EL}^{++}$  ontologies on the web is selected, the test suite comprises SUMO<sup>14</sup>, NCI<sup>15</sup>, and GO<sup>16</sup>. Information about those ontologies is described in Table 2.

<sup>9</sup> <http://pellet.owldl.com/>

<sup>10</sup> <http://owlapi.sourceforge.net/>

<sup>11</sup> For nominal  $\{a\}$ , we deal with it as concept to test the relevance.

<sup>12</sup> It uses DefaultExplanationGenerator in Pellet to get the  $mMod$ , we implement  $eMod$  extraction based on  $mMod$  extraction.

<sup>13</sup>  $LoMod$  is  $\perp$ -module and  $UpMod$  is  $\top$ -module, they can be obtained from <http://krono.act.uji.es/people/Ernesto/safety-ontology-reuse>

<sup>14</sup> <http://www.ontologyportal.org/>

<sup>15</sup> <http://www.mindswap.org/2003/CancerOntology/>

<sup>16</sup> <http://www.geneontology.org/index.shtml>

**Table 2.** Information about ontologies

ontology	expressivity	#concepts	#objectProperties	#individuals
SUMO	RDFS	630	217	435
NCI	$\mathcal{EL}^{++}$	27652	70	0
GO	$\mathcal{EL}^{++}$	26231	5	154932

We generate axioms which contain only useful symbols from the above ontologies. The generated axioms are  $C \sqsubseteq D$  type,  $C$  and  $D$  are concepts. To assure  $C, D$  are useful symbols in the axiom  $C \sqsubseteq D$ ,  $C$  should not be the bottom concept  $\perp$  and  $D$  should not be the top concept  $\top$  in the ontology. Those conditions can be easily guaranteed. All ontology classification axioms in SUMO is collected for module extraction<sup>17</sup>. As NCI and GO contain too many  $C \sqsubseteq D$  axioms, we randomly generate 5000 entailed axioms ( $\mathcal{O} \models C \sqsubseteq D$  and  $C \sqsubseteq D \notin \mathcal{O}$ ) and 500000 not entailed axioms ( $\mathcal{O} \not\models C \sqsubseteq D$ ) for evaluation. Entailment information for them is described in the second and third column in Table 3.

Boundary-based module is firstly compared with the optimal module: minimal module and essential module, and the evaluation is described in Table 3. Information of minimal boundary-based module is illustrated in the fourth and fifth column. It includes number of not entailed axioms whose  $bMod$  is not empty, which means it is not equal to  $mMod$  and  $eMod$ , plus the average size of those module; and average size of  $bMod$  for entailed axioms. Average size of  $eMod$  and  $mMod$  for entailed axioms are described in the last two columns.

**Table 3.** Comparison information between  $bMod$  and optimal module on  $\mathcal{O} \models C \sqsubseteq D$ 

Ont.	entailment Inf.		$bMod$		$eMod$	$mMod$
	Num of $\mathcal{O} \models \alpha$	Num of $\mathcal{O} \not\models \alpha$	$\mathcal{O} \not\models \alpha$ nonempty (Num./Avg. Size)	Avg. Size for $\mathcal{O} \models \alpha$	Avg. Size for $\mathcal{O} \models \alpha$	Avg. Size for $\mathcal{O} \models \alpha$
SUMO	3462	392112	0	4.92	4.92	4.21
NCI	5000	500000	169(0.034%)/15.50	10.43	8.31	4.23
GO	5000	500000	72(0.015%)/15.63	12.03	8.60	3.98

From Table 3, we can clearly see that  $bMod$  is very close to the  $eMod$ . For not entailed axioms, all  $bMod$  are empty in SUMO ontology, in other words, they are equal to the  $mMod$  and  $eMod$ ; in NCI and GO ontology, there are only 0.034% and 0.015% axioms whose module is not empty, the average size is 15.50 and 15.63. For entailed axioms, in SUMO ontology, the average size of  $bMod$ ,  $eMod$  and  $mMod$  are 4.92, 4.92 and 4.21; in NCI ontology, they are 10.43, 8.31 and 4.32; in GO ontology, they are 12.03, 8.60 and 3.98.

It is easy to observe that  $bMod$  is almost optimal for not entailed axioms,  $bMod$  which is not empty only account for a very small proportion. For entailed axioms,  $bMod$

<sup>17</sup> Axioms which are already included in  $\mathcal{O}$  are eliminated.

is very close to the essential module. As our minimal minimal boundary-based module is an approximated approach for essential axiom-based module, when essential module is much bigger than minimal module, the  $bMod$  is not very close to the minimal module, however it is small enough for optimizing reasoning.

We also compute the  $LoMod$  and  $UpMod$  for the above generated axioms. For  $\mathcal{O} \models C \sqsubseteq D$ , the  $LoMod(C)$  and  $UpMod(D)$  also exact axiom-based module [9]. The comparison information is illustrated in Table 4. For each method, we calculate the average and the maximal size of these modules.

**Table 4.** Comparison information between  $bMod$  and locality module on  $\mathcal{O} \models C \sqsubseteq D$

ontology	$LoMod$		$UpMod$		$bMod$	
	Avg. Size	Max Size	Avg. Size	Max Size	Avg. Size <sup>18</sup>	Max Size
SUMO	45.37	1751	764.54	778	2.45	19
NCI	2290.65	29252	36.97	437	9.09	53
GO	32503.28	104883	154953	155151	9.92	69

From Table 4, it is easy to observe that the locality-based module contains too much unrelated axioms for reasoning task, and our boundary-based module is more suitable for optimizing ontology reasoning.

## 7 Conclusion and Future Work

In this paper, we described an axiom-based module by analyzing the relationship between axioms and boundaries of interpretations of symbols. We proved that the boundary-based module is exact for reasoning tasks. As most work of module extraction can be done offline, the boundary-based module extraction can have good run-time performance. Experiments show that minimal boundary-based module is very close to the optimal module, and is suitable for optimizing ontology reasoning.

In the future, we will extend our module extraction method to test more complex ontologies such as GALEN and SWEET. Next, we plan to apply the boundary-based modularization in improving the performance of an ontology reasoner, e.g., Pellet. Finally, other ontology reasoning task such as incremental reasoning and analogical reasoning by using boundary-based module will be also investigated.

## Acknowledgement

We are very grateful to Ernesto Jiménez-Ruiz and Bernardo Cuenca Grau for providing the locality-based module extraction source code. This work is supported in part by the National Natural Science Foundation of China under Grant No.60675015.

<sup>18</sup> We assume that entailed axioms, not entailed axioms whose  $bMod$  is empty and not entailed axioms whose  $bMod$  is nonempty each accounts for 50%, 25% and 25% in total reasoning axioms. Note that it is a conservative assumption, from the Table 3, we can see that not entailed axioms whose  $bMod$  is empty account for very high proportion in all not entailed axioms.

## References

- [1] Serafini, L., Borgida, A., Tamilin, A.: Aspects of distributed and modular ontology reasoning. In: IJCAI. (2005) 570–575
- [2] Grau, B.C., Parsia, B., Sirin, E.: Working with multiple ontologies on the semantic web. In: International Semantic Web Conference. (2004) 620–634
- [3] Bao, J., Caragea, D., Honavar, V.: A tableau-based federated reasoning algorithm for modular ontologies. In: IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Press (2006) 404–410
- [4] Cuenca Grau, B., Halaschek-Wiener, C., Kazakov, Y.: History matters: Incremental ontology reasoning using modules. In: Proceedings of the 6th International Semantic Web Conference. Volume 4825., Springer-LNCS (2007) 183–196
- [5] Bao, J., Honavar, V.: Divide and conquer semantic web with modular ontologies - a brief review of modular ontology language proposals. In: ISWC 2006 Workshop on Modular Ontologies (WoMo 2006). (2006)
- [6] Stuckenschmidt, H., Klein, M.: Structure-based partitioning of large concept hierarchies. In: Proceedings of the 3th International Semantic Web Conference. Volume 3298., Springer-LNCS (2004) 289–303
- [7] Seidenberg, J., Rector, A.: Web ontology segmentation: analysis, classification and use. In: Proceedings of the 15th international conference on World Wide Web, New York, NY, USA, ACM (2006) 13–22
- [8] Noy, N.F., Musen, M.A.: Specifying ontology views by traversal. In: Proceedings of the 3th International Semantic Web Conference. Volume 3298., Springer-LNCS (2004) 713–725
- [9] Jimenez-Ruiz, E., Cuenca Grau, B., Schneider, T., Sattler, U., Berlanga, R.: Safe and economic re-use of ontologies: a logic-based methodology and tool support. In: Proceedings of the 5th European Semantic Web Conference. Volume 5021., Springer LNCS (2008) 185–199
- [10] Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Modular reuse of ontologies: Theory and practice. *Journal of Artificial Intelligence Research* **31** (2008) 273–318
- [11] Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Just the right amount: Extracting modules from ontologies. In: Proceedings of the 16th International World Wide Web Conference, Springer-LNCS (2007) 717–727
- [12] Seidenberg, J., Rector, A.L.: Web ontology segmentation: analysis, classification and use. In: WWW. (2006) 13–22
- [13] D’Aquin, M., Schlicht, A., Stuckenschmidt, H., Sabou, M.: Ontology modularization for knowledge selection: Experiments and evaluations. In: DEXA. (2007)
- [14] Baader, F., Brandt, S., Lutz, C.: Pushing the envelope further. In Clark, K., Patel-Schneider, P.F., eds.: In Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions. (2008)
- [15] Suntisrivaraporn, B.: Module extraction and incremental classification: A pragmatic approach for  $\mathcal{EL}^+$  ontologies. In: Proceedings of the 5th European Semantic Web Conference. Volume 5021., Springer-LNCS (2008) 230–244
- [16] Tsarkov, D., Horrocks, I.: DL reasoner vs. first-order prover. In: Description Logics. Volume 81., CEUR-WS (2003) 152–159
- [17] Kalyanpur, A., Parsia, B., Horridge, M., Sirin, E.: Finding all justifications of owl dl entailments. In: Proceedings of the 6th International Semantic Web Conference. Volume 4825., Springer-LNCS (2007) 267–280
- [18] 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, USA (2003)