A METHOD FOR APPROXIMATION TO ONTOLOGY REUSE PROBLEM

Eleni Tsalapati, Giorgos Stamou and Giorgos Koletsos

School of Electrical and Computer Engineering, National Technical University of Athens
GR-15773 Zographou, Athens, Greece

Keywords: Ontology Reuse, Module, Locality.

Abstract: We describe a methodology that approximates the safety condition in the integration of ontologies. Intuitively, the safety condition ensures that, after the reuse, the external ontology remains intact. It was recently proved that it is a problem of high complexity to examine whether this condition holds. In order to approximate the problem, we suggest a method by which the ontology designer that reuses concepts from an ontology network will be informed about the degree in which the safety is violated and according to this information she can choose the appropriate ontologies to reuse concepts from.

1 INTRODUCTION

Today, one can find in the Web a vast number of ontologies on the same domain in different levels of granularity or in different domains, available for reuse. The research conducted on the problem of ontology reuse process, regardless of the approach method, has revealed some very crucial obstacles. It is essential that the ontology integration should be carried out in such a way that the consequences of the axioms of the external ontology are not changed when its elements are reused in the ontology under development. It is proved (Lutz et al., 2007) that to question whether this condition holds, even in languages of very low expressivity, is practically undecidable (2EXPTIME-complete).

In this paper, we assume that the ontology designer wishes to generalise (e.g., add new superconcepts) some of the concepts of the ontology under development by importing concepts from an external one. At the same time though, we do not prevent the developer from refining (e.g., add new subconcepts) some concepts, as well. However, the simultaneous generalization and refinement from the same external ontology may introduce new knowledge to this ontology. For instance, if the ontology developer reuses in his ontology P the concepts X, Y from the ontology Q to create the following axioms: $A \sqsubseteq Q : X$ and $Q : Y \sqsubseteq B$, and the axiom $B \sqsubseteq A$ is already present in P, then the axiom $Q : Y \sqsubseteq Q : X$ is implied, which may not hold in Q. In the same way, non anticipated

changes may also occur in the ontology P.

We have developed a tool that provides the ontology developer with an estimation of the changes caused not only in the external ontology, but also in the ontology under development. Being aware of the effects of the integration of the two ontologies, the user is enabled to choose firstly, whether he will refine a concept or not, and secondly, the most appropriate external ontology to choose concepts from, based on the criterion "to cause the least possible damage".

2 PRELIMINARIES

We consider ontologies to be a set of axioms in a descriptions logic language $\mathcal{L}$. Description Logic is the language that underlies the modern ontology languages, like OWL DL. The signature of an ontology O is the set of atomic concepts (**C**), atomic roles (**R**) and individuals (**I**) that occur in O. An ontology based on $\mathcal{SHOIQ}$ language contains also nominals.

The notions of logical entailment, S-interpretation, restriction of an S-interpretation and satisfiability follow the definitions of the D. L. Handbook (Baader et al., 2003). We omit their description in this paper due to limited space. An expansion of an S_1-interpretation I_1 to S is an S-interpretation I such that $I|_{S_1} = I_1$ (Grau et al., 2007b). In bottom expansion, atomic concepts and roles, not in S are interpreted as Bottom ($\perp$), whereas

in top expansion, they are interpreted as Top ($\top$).

The reuse of an external ontology can lead to undesirable side effects, such as the entailment of new axioms or even inconsistencies involving the reused vocabulary. However, a *safe* ontology for a signature S, enables the implementation of the ontology reuse process without any fallacies:

Definition 2.1. (Safety for a Signature, (Grau et al., 2007b)). *Let $\mathcal{L}$ be an ontology language, and let P be an ontology and S a signature over $\mathcal{L}$. We say that P is safe for S w.r.t. $\mathcal{L}$, if for every ontology Q over $\mathcal{L}$ with $Sig(P) \cap Sig(Q) \subseteq S$, we have that $P \cup Q$ is a conservative extension of Q w.r.t. $\mathcal{L}$.*

The conservative extension property ensures that the additional axioms of the ontology under development will not add any new consequences over the vocabulary of the reused ontology.

Following, we describe the syntactic definitions $\top$-locality and $\perp$-locality concepts (Jiménez-Ruiz et al., 2008), as they are heavily used in our method.

Definition 2.2. (Syntactic $\perp$-Locality and $\top$-Locality for *SHIQ*). *Let S be a signature. An axiom a is $\perp$-local w.r.t. S ($\top$-local w.r.t S) if $a \in Ax(S)$, as defined in (a) ((b)), where C and R denote arbitrary concepts and roles. An ontology O is $\perp$-local ($\top$-local) w.r.t. S if a is $\perp$-local ($\top$-local) w.r.t. S for all $a \in O$.*

(a) $\perp - \textbf{Locality}$. *Let $A^{\perp}, R^{\perp} \not\in S$, $C^{\perp} \in \textbf{Bot(S)}$, $C^{\top}_{(i)} \in \textbf{Top(S)}$, $n \in \textbf{N} \setminus \{0\}$:*

$\textbf{Bot(S)} ::= A^{\perp} \mid \perp \mid (\neg C^{\top}) \mid (C \sqcap C^{\perp}) \mid (C^{\perp} \sqcap \sqcap C) \mid (\exists R^{\perp}.C) \mid (\exists R.C^{\perp}) \mid (\geq nR^{\perp}.C) \mid (\geq nR.C^{\perp})$

$\textbf{Top(S)} ::= \top \mid (\neg C^{\perp}) \mid (C^{\top}_1 \sqcap C^{\top}_2) \mid \geq 0R.C$

$\textbf{Ax(S)} ::= C^{\top} \sqsubseteq C \mid C \sqsubseteq C^{\perp} \mid R^{\perp} \sqsubseteq R \mid Trans(R^{\perp})$

(b) $\top - \textbf{Locality}$. *Let $A^{\top}, R^{\top} \not\in S$, $C^{\perp} \in \textbf{Bot(S)}$, $C^{\top}_{(i)} \in \textbf{Top(S)}$, $n \in \textbf{N} \setminus \{0\}$:*

$\textbf{Bot(S)} ::= (\neg C^{\top}) \mid \perp \mid (C \sqcap C^{\perp}) \mid (C^{\perp} \sqcap \sqcap C) \mid (\exists R.C^{\perp}) \mid (\geq nR.C^{\perp})$

$\textbf{Top(S)} ::= A^{\top} \mid \top \mid (\neg C^{\perp}) \mid (C^{\top}_1 \sqcap C^{\top}_2) \mid (\exists R^{\top}.C^{\top}) \mid (\geq nR^{\top}.C^{\top}) \mid \geq 0R.C$

$\textbf{Ax(S)} ::= C^{\top} \sqsubseteq C \mid C \sqsubseteq C^{\perp} \mid R \sqsubseteq R^{\top} \mid Trans(R^{\top})$

If an ontology P is $\perp$-local or $\top$-local w.r.t. a signature S, then P is safe for S (Grau et al., 2007a). The simultaneous refinement and generalization of an ontology from the same external one, may compromise safety (Jiménez-Ruiz et al., 2008). Particularly, when the signature S is to be refined (generalized), the ontology P must be $\perp$-local ($\top$-local) w.r.t. S. Consequently, in the case of refinement of the signature S (and thus generalization of P) the set of non-$\perp$-local axioms contains the axioms that threaten the safety.

3 METHOD FOR PARTIAL ONTOLOGY REUSE

The goal of this method is to enable the ontology developer to refine and generalise several concepts from the same external ontology by providing him with an estimation of the non-safety degree resulted from this ontology reuse. Particularly, we are interested in providing the user not only with the affected part of the external ontology Q, but also with the non anticipated changes (new axioms over the language of P) caused in the ontology under development.

According to the previous paragraph the set of the non-$\perp$-local axioms w.r.t. $Sig(Q)$ equals to a superset of the axioms that threaten the safety of the integration. To clarify this, we demonstrate the following example, in which two ontologies developed for the BOEMIE Project (www.boemie.org), the Athletic Event Ontology, aeo, and the Geographic Information Ontology, gio are used.

Example 3.1. *Supposing that SportsPOI is concept of aeo ontology and aeo reuses concepts from gio, then the following is a non-$\perp$-local axiom wrt $Sig(gio)$:*

$$gio:Stadium \sqsubseteq SportsPOI \tag{1}$$

If the following axioms also exist in aeo ontology :

$$gio:Stadium \sqsubseteq \exists isLocatedIn.City \sqcap \sqcap SportsPOI \sqcap \exists hasArea.TrackArea \tag{2}$$

$$\exists isLocatedIn.City \sqcap SportsPOI \sqsubseteq gio:ManMadeFeature \tag{3}$$

Then following new axiom of gio is implied:

$$gio:Stadium \sqsubseteq gio:ManMadeFeature \tag{4}$$

Apart from the fact that this axiom might introduce an inconsistency in gio, it also adds a new definition on the concept Stadium, that did not initially exist. We call the axiom (1) as directly affected axiom.

The set of *directly affected axioms* is the remaining set of non-$\perp$-local axioms after the abstraction of the axioms already included in gio:

Definition 3.2. *Let $\mathcal{L}$ be an ontology language and P and Q be ontologies over $\mathcal{L}$. If ontology P reuses concepts from ontology Q, then the set A of the directly affected axioms of P for Q is:*

$$A = \{a \in P \mid a : non - \perp - localAxiom \, w.r.t. \, Sig(Q) \wedge \\ \wedge a \not\in Q\}$$

In Example 3.1 the definition of the concept Stadium of gio ontology, is changed. To approximate the subset of the reused entities *affected* by this integration it suffices to find the subsumed concepts of the set of directly affected axioms. For instance, in Example 3.1, the Stadium concept is a *directly affected entity*. At

present this theory concerns only simple axioms (axioms of the form $A \sqsubseteq C$, where A is atomic concept). The set of directly affected entities of an axiom a by an ontology P is symbolized as $DAE(a,P)$. The set of directly affected entities of an ontology Q by an ontology P is symbolized as $DAE(Q,P)$ and represents the set of affected entities of each axiom of Q.

The modification of the Stadium concept definition also affects all the concepts that are related directly or indirectly to Stadium in gio. To make this more clear we demonstrate the following example:

Example 3.3. *Suppose that gio contains the following axioms:*

$$IndoorStadium \sqsubseteq Stadium \qquad (5)$$

$$OutdoorStadium \sqsubseteq Stadium \qquad (6)$$

then the concepts IndoorStadium and OutdoorStadium are also altered. The modification of these definitions is propagated to other concepts related to them. For instance, the definition of IndoorAthleticCenter concept is also affected:

$$\begin{aligned} IndoorAthleticCenter \sqsubseteq \\ \exists contains.IndoorStadium \end{aligned} \qquad (7)$$

Thus, to find the total set of the affected concepts is suffices to find all the subconcepts, directly and indirectly related to an affected entity. To achieve this we calculate, the set of $\top$-local axioms of ontology Q w.r.t. $S' = Sig(Q) \setminus S$, where S is the set of affected entities of Q by ontology P. Intuitively, in this way, all affected entities become equivalent to the top concept and thus, the set of $\top$-local axioms includes the subsumption axioms that their right hand side concept is affected. If an axiom is found to be $\top$-local then it is added to the set of affected axioms and its directly affected entities are abstracted from S', thus the next checked axiom is tested w.r.t. $\top$-locality for the new S'. More formally, the set of *affected axioms* is defined as:

Definition 3.4. *Let P and Q be ontologies, such that P reuses concepts from Q. Then the set of affected axioms of Q by ontology P is defined as:*

$$\begin{aligned} AA(Q,P) = \{a \in Q | a : \top - local \, w.r.t. \, (Sig(Q) \setminus \\ (DAE(Q,P) \cup DAE(Q,AA(Q,P))))\} \end{aligned}$$

Similarly, if P is the ontology under development and P' the ontology P before the reuse of the external ontology, the set of affected axioms of the ontology P' is $AA(P',P)$. To calculate the *rate* of affected axioms of an ontology Q from an ontology P, $r(Q,P)$, it suffices to divide the number of axioms of the set $AA(Q,P)$ to the total number of axioms of Q (i.e. $r(Q,P) = |AA(Q,P)| / |Q|$).

Having the rates $r(Q,P)$ and $r(P',P)$, we can calculate the total cost of the reuse process in relation to two weights w_1 and w_2, that determine how important is the particular change to each ontology:

Definition 3.5. (Total Cost) *Let P and Q be ontologies, S the set of entities of Q reused by P and w_1, $w_2 > 0$ the weights. The total cost of the ontology reuse procedure is:*

$$TotalCost = \frac{w_1 \cdot r(Q,P) + w_2 \cdot r(P',P)}{w_1 + w_2}$$

Following, we demonstrate the algorithm that computes the affected axioms of ontology Q from P.

Algorithm 1: AA(Q,DAE(Q,P)).

Input:
$DAE(Q,P)$:set of Directly Affected Entities of Q from P
Q: external ontology
Output:
Q_2: Affected Axioms

```
1:  S ← (Sig(Q) \ DAE(Q,P)), Q₁ ← ∅, Q₂ ← Q
2:  while  (¬empty (Q₂)){
3:    forall  a ∈ {a|a ∈ Q₂ ∧ a : OWLSubClassAxiom} {
4:      if  (⊤ − locality − test(a,S)){
5:        Q₁ = Q₁ ∪ {a}
6:        S = S \ {LHS(a)}
7:        Q₂ = Q \ Q₁}
8:      else  {Q₂ = Q₂ \ {a}}
9:  }}
10: return  Q₂
```

For the locality-test we used the method provided by PRoSE (http://krono.act.uji.es/people/Ernesto/safety-ontology-reuse/proSE-current-version). The LHS function returns the left hand side of a subclass axiom.

Proposition 3.6. *For every input P and Q, the Algorithm 1 computes the affected axioms of the participating ontologies.*

Proof (Sketch). The algorithm terminates as in every iteration of the while loop the size of Q_2 decreases. Initially, all the $\top$-local axiom are transferred from Q_2 to Q_1, and then the rest axioms, which are non-$\top$-local, are removed from Q_2. Note, that the syntactic locality test has polynomial complexity.

The output $AA(Q,DAE(Q,P))$ is the set of affected axioms of Q from P, as every axiom that is neither in Q_1 nor in Q_2 is non-$\top$-local wrt $Sig(Q) \setminus (DAE(Q,P) \cup DAE(Q,AA(Q,P)))$, since, as indicated in line:8, it remains in the set $Q \setminus (Q_1 \cup Q_2)$. □

In order to evaluate this algorithm, we tested it with the ontologies aeo and gio:

Example 3.7. *After the ontology reuse, aeo contains 2473 axioms and gio 607. The current form of aeo*

ontology, has generalized only some of its concepts by reusing external concepts from gio ontology. Thus, gio ontology is not threatened by aeo. For the purpose of this example, we added the following axioms to aeo:

Table 1: Experimental results.

Axiom	Thr. Ax. %	Time
g:GeographicObject ⊑ HLC	33.4	7.06
g:Polygon ⊑ HLC	0	6.87
g:Surface ⊑ HLC	0.4	9.51
g:Building ⊑ ∃hasPart. MLC	2.6	7.79

The first column of the table demonstrates the axioms added, the second the percentage of gio ontology threatened by the corresponding axiom and the third the time (seconds) required to return the results.

The GeographicObject concept of gio ontology is relatively high in the hierarchy and thus the potential modification of its definition will affect a large part of gio ontology. On the other hand, Polygon concept is on the bottom of the ontology and thus does not affect any part of gio ontology. Consequently, the user is allowed to refine this concept, although other concepts of gio have been generalized, without any change on the information provided by gio. The Surface concept affects the following axioms:

$$Polygon \sqsubseteq Surface \qquad (8)$$

$$MultiSurface \sqsubseteq \exists hasGeometryObject. \, Surface \quad (9)$$

$$MultiPolygon \sqsubseteq MultiSurface \qquad (10)$$

which were traced by our method (0,4% of the whole ontology). The modification of the Building concept threatens 16 other subconcepts of this concept.

4 RELATED WORK

Two basic directions define the research areas developed in the field of knowledge reuse. The first direction is concentrated on the safe ontology reuse. The second direction is focused on the development of appropriate formalisms that enable the ontology reuse via conceptualized interpretations (Bao et al., 2007).

The research based on the first direction has lead to the exploitation of the notion of module and the development of tools like PRoSE, which enables the safe ontology reuse but under very strict constraints (the user is allowed only to generalise or to refine the concepts of the target ontology). Another interesting approach is suggested by (Suntisrivaraporn, 2008), where a new kind of module is introduced, but it concerns only $\mathcal{EL}+$ ontologies.

5 CONCLUSIONS - FURTHER RESEARCH

In this study we described a method by which we can approximate the ontology reuse problem. After the integration of two ontologies, the ontology designer is informed about the extent of the loss of information that will occur in order to maintain the safety state. In this way, the designer can choose the appropriate ontology to reuse concepts from, based on the criterion of causing the least loss of information.

For further research, we are interested in adjusting the method in order to deal with more complex axioms than the simple axioms. We are also interested in examining the condition of safety, in case the set of refined entities belongs to a disjoint module from the set of generalised entities. Finally, another direction followed by our research could be to investigate the ontology reuse problem in the case of interaction of more than two ontologies.

REFERENCES

Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F., editors (2003). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.

Bao, J., Slutzki, G., and Honavar, V. (2007). A semantic importing approach to knowledge reuse from multiple ontologies. In *AAAI*, pages 1304–1309.

Grau, B. C., Horrocks, I., Kazakov, Y., and Sattler, U. (2007a). Extracting modules from ontologies: Theory and practice. Technical report, The University of Manchester.

Grau, B. C., Horrocks, I., Kazakov, Y., and Sattler, U. (2007b). Just the right amount: extracting modules from ontologies. In *WWW*, pages 717–726.

Jiménez-Ruiz, E., Grau, B. C., Sattler, U., Schneider, T., and Llavori, R. B. (2008). Safe and economic re-use of ontologies: A logic-based methodology and tool support. In *ESWC*, pages 185–199.

Lutz, C., Walther, D., and Wolter, F. (2007). Conservative extensions in expressive description logics. In *IJCAI*, pages 453–458.

Suntisrivaraporn, B. (2008). Module extraction and incremental classification: A pragmatic approach for el+ ontologies. In *ESWC*, LNCS. Springer Verlag.