A Proposal for Screening Inconsistencies in Ontologies based on Query Languages using WSD

Chieko NAKABASAMI
Toyo University
1-1-1 Izumino, Itakura,
Oura, Gunma, Japan, 374-0193
chiekon@itakura.toyo.ac.jp

Naoyuki NOMURA
Hosei University
2-17-1 Fujimi, Chiyoda-ku,
Tokyo, Japan, 102-8160
nomuran@tokyo.email.ne.jp

Abstract

In this paper, we discuss a method to screen inconsistencies in ontologies by applying a natural language processing (NLP) technique, especially, those used for word sense disambiguation (WSD). In the database research field, it is claimed that queries over target ontologies should play a significant role because they represent every aspect of the terms described in each ontology. According to (Calvanese et al., 2001), considering the global and the local ontologies, the terms in the global ontology can be viewed as the query over the local ontology, and the mapping between the global and the local ontologies is given by associating each term in the global ontology with a view. On the other hand, ontology screening systems should be able to take advantage of some popular techniques for WSD, which is supposed to decide the right sense where the target word is used in a specific context. We present several examples regarding inconsistencies in ontologies with the aid of DAML+OIL notation(DAML+OIL, 2001), and propose that WSD can be one of the promising method to screen such as inconsistencies.

1 Introduction

In recent years, the semantic web (Berners-Lee et al., 2001) has been evolving as the next-generation web technology and has attracted the attention of many researchers in database and knowledge engineering communities. In the semantic web, contributions obtained from fields related to databases frequently refer to ontology maintenance, reuse, and sharing. Based on the results from database research, this paper proposes a method to screen inconsistencies in ontologies by applying a natural language processing (NLP) technique, especially, those used for word sense disambiguation (WSD).

Many reference books on WSD are available (e.g., (Manning and Schutze, 1999)). As for ontology integration, several approarches are proposed (Mitra et al., 2001)(Euzenat, 2001). In (Calvanese et al., 2001), global-centric (aka global-as-view) and local-centric (aka local-as-view) approaches for an ontology integration framework are proposed, respectively. In this paper, we support the global-as-view approach for screening inconsistencies in ontology. In addition, (Calvanese et al., 2001) claims that queries over target ontologies should play a significant role because they represent every aspect of the terms described in each ontology. The terms in the global ontology can be viewed as the query over the local ontology, which describes a concept definition. On the other hand, ontology screening systems should be able to take advantage of some popular techniques for WSD, which is supposed to decide the right sense where the target word is used in a specific context. We claim that WSD is a promising method for determining which local ontology should be used for forming concepts for a global ontology. This paper contains a brief introduction of the global-centric approach, which is described in Section 2. In addition, we mention the inconsistencies in ontologies caused by words with multiple definitions. For example, the word gbassh is chosen, and we explain that some queries based on two of its definitions have concepts in the global ontology. After extracting new concepts by querying the global ontology, each local ontology is illustrated with DAML+OIL (DAML+OIL, 2001). Several inconsistencies in the ontology are presented in Section 3 over the global ontology, and the WSD for solving such inconsistencies are discussed. The final section is the conclusion.

2 Global-as-view approach

In (Calvanese et al., 2001), global-as-view and local-as-view approaches are proposed for ontology integration. According to the global-as-view approach, the mapping between the global and the local ontologies is given by associating each term in the global ontology with a view. Let C be a term in the global ontology G, V a query language over the terms of the local ontologies, and M the mapping between the global and the local ontologies. Given that D is a local model for the ontology integration system and I a global interpretation for the system, the correspondence between C and V is specified as follows by referring to (Calvanese et al., 2001):

- $\langle C, V, sound \rangle$

  if all tuples satisfying V in D satisfy C in I

- $\langle C, V, complete \rangle$

  if no tuple other than those satisfying V in D satisfies C in I

- $\langle C, V, exact \rangle$

  if the set of tuples that satisfy C in I is exactly the set of tuples satisfying V in D.

In the above notation, "I satisfies" means that I satisfies every correspondence in the mapping between the global ontology and the local ontologies wrt D. These correspondences are valid if the global ontology is assumed to be consistent; however, inconsistencies might occur when the term in the global ontology has more than one definition. We use the word "*bass*," for example, which has at least two definitions, one, "*a man whose singing voice is very low*," and another, "*a kind of fish*" (LDOCE, 1995). Each definition is represented in order with first-order-language-like notation, as follows:

$$C(x) \leftarrow singer(x), voice(x, low).$$
$$C(x) \leftarrow aKindOf(x, fish).$$

Some concepts in the global ontology are described using the above two $C(x)$s. The following concept is represented with the first $C(x)$ because $C_1(x)$ is a member of an orchestra, which would be impossible that $C_1(x)$ is a fish from the standard context.

$$C_1(x) \leftarrow C(x), isMemberOf(x, orchestra). \quad (1)$$

The concept $C(x)$ in (1) can be represented with the DAML+OIL notation, as shown below. This concept is assumed to belong to one local ontology (for example, *local-ont-1*). In the representation, some parts in the representation, such as namespace prefixes and URIs, are omitted because they do not directly relate to this paperfs intention. In addition, the referenced resources and properties whose definitions are not in the representation are supposed to be defined implicitly.

```
<rdf:RDF>
<daml:Ontology rdf:about=" ">
<rdfs:comment>local-ont-1
</rdfs:comment>
</daml:Ontology>
<daml:Class rdf:ID="bass">
<rdfs:subClassOf rdf:resource="#singer"/>
<hasVoiceOf rdf:resource=#very_low />
</daml:Class>
<daml:ObjectProperty rdf:ID="hasVoiceOf">
<rdfs:range rdf:resource="#voice"/>
</daml:ObjectProperty>
<daml:Class rdf:ID="voice">
<daml:oneOf rdf:parseType="daml:collection">
<voice rdf:ID="high/">
<voice rdf:ID="medium"/>
<voice rdf:ID="low"/>
<voice rdf:ID="very_low"/>
</daml:oneOf>
</daml:Class>
</rdf:RDF>
```

$C_1(x)$ is assumed to be a result from a query over the global ontology, "$C(x), memberOf(x, orchestra)$" The set of extensions of C_1 is returned as a result of the query. Once C_1 is formed, 'x' as a result of the global ontology becomes a set of instances for a new class, and the local ontology is revised by adding a subclass of the existing class. If such a new subclass were named "*sub-bass*," it would be represented in this manner:

```
<daml:Class rdf:ID="sub-bass">
<subClassOf rdf:resource="#bass"/>
</daml:Class>
```

```
<daml:Class rdf:about="#sub-bass">
<isMemberOf rdf:resource="#orchestra"/>
</daml:Class>
```

Let O be an ontology screening system. In this case, O obtains one concept "*sub-bass,*" i.e., "*a bass singer working at an orchestra,*" from the intersection between $C(x)$ and $isMemberOf(x, orchestra)$. The concept "*sub-bass*" is formed by combining the local and global ontologies. Considering this example, appropriate extensions of the intended concept in the global ontology are obtained and recognized only if the right sense of the concept in the local ontology is used. A different concept might be drawn from different senses from other local ontologies, but the high rate of co-occurrence between singer and orchestra led by WSD would play a crucial role in preventing another sense from being considered. Let's consider another example:

$$C_2(x) \leftarrow C(x), livesIn(x, river). \qquad (2)$$

In (2), x is likely the second sense of "*bass.*" For example, x is an extension of the bass class as a subclass of fish. WSD would justify it because the co-occurrence of "*fish*" and "*river*" is rather strong. The local ontology of "*bass*" from the second sense (*local-ont-2*) is represented with DAML+OIL, as shown below. It is also revised when a new subclass from query (2) is added.

```
<rdf:RDF>
<daml:Ontology rdf:about=" ">
<rdfs:comment>local-ont-2
</rdfs:comment>
</daml:Ontology>
<daml:Class rdf:ID="bass">
<rdfs:subClassOf rdf:resource="#fish"/>
</daml:Class>
<daml:Class rdf:ID="sub-bass">
<subClassOf rdf:resource="#bass"/>
</daml:Class>
<daml:Class rdf:about="#sub-bass">
<livesIn rdf:resource="#river"/>
</daml:Class>
</rdf:RDF>
```

3 Screening inconsistencies in the global ontology

According to (Calvanese et al., 2001), the following situations involving inconsistency or ambiguity occur as a result of the mapping of global and local ontologies:

1. There are no global models.

   This occurs when the data in the extension of the local ontologies do not satisfy the constraints for functional attributes. For example, in (Calvanese et al., 2001), a property "*age*" is considered. Since "*age*" is a function, the value as its range must be defined with only one value. However, the global ontology does not have a model concerning "*age*" any more if "*age*" has more than one value from the constraint of the concepts in a local ontology.

2. There are several global models.

   This occurs when the data in the extension of the local ontologies do not satisfy the ISA relationships of the global ontology: i.e., several ways exist to add suitable instances to the elements of the global ontology in order to satisfy the constraints from a local ontology. For example, it is assumed that a "*student*" must be enrolled in a certain university. More than one university can be considered when the constraints in a local ontology do not mention which university each extension of "*student*" is enrolled in from the concept definitions of the local ontology. Such ambiguities depend on how precisely the concepts of a local ontology are defined. As a result, all interpretations for an intended ontology integration system within which a valid concept exists must be accepted as models because of the potential ambiguities present in local ontologies.

We propose two other solutions to the unsuitable situations described above; the first relates to yielding invalid extensions, and the second to forming wrong concept definitions as a result of WSD failure. For an explanation of the first situation, consider this example:

Example 1. Re-considering (2) in Section 2.
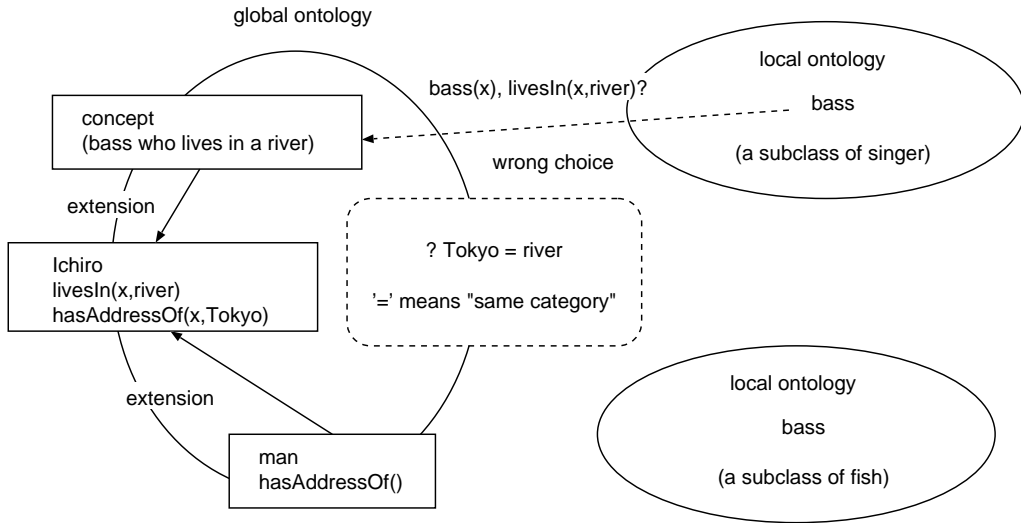
$$C_2(x) \leftarrow C(x), livesIn(x, river). \qquad (3)$$

Figure 1: An inconsistency resulting from a wrong choice from word "*bass*"

In (3), concept C_2 is assumed to be formed from concept C, which has the meaning of "*bass as a singer*." The meaning of C_2 is "*a bass that lives in a certain river*." Provided that "*Ichiro*" is an extension of C_2, it could be said that *Ichiro* lives in a river. In an ontology hierarchy, there would be multiple inheritances around one term. Suppose "*Ichiro*" is also an extension of the class "*man*." If the class "*man*" has a property called "*isAddressOf*," which is defined as the same property of "*livesIn*," "*Ichiro*" must have an instantiated address. When his address is "*Tokyo*," an inconsistency arises because "*Tokyo*" and "*river*" are not classified in the same category. In addition, this inconsistency in terms of having an inappropriate value of properties is propagated towards lower subclasses of the target class. This situation is expressed in Figure 1. We claim that a situation such as this should be solved using WSD because of the strong relationship between "*fish*" and "*river*" mentioned in Section 2.

Example 2. The second situation concerns WSD failure. Suppose the following query:

$$C_3(x) \leftarrow C(x), hasHobbyOf(x, swimming). \tag{4}$$

In (4), C_3 would be assumed to be a concept referring to a subclass of person from the ordinary context. However, the ontology screening system might classify the concept C_3 into the wrong place in the global ontology if it takes the concept C as the second meaning of "*bass*," i.e., a kind of fish. Deriving the related concepts in the global ontology on the basis of a wrong concept whose meaning is never intersectable with the right one causes an unexpected inconsistency over the global ontology. After C_3 is mistaken, the local ontology would result in a concept with the meaning "*a bass whose hobby is swimming*" in spite of the fact that a bass is a type of fish.

```
<rdf:RDF>
<daml:Ontology rdf:about=" ">
<rdfs:comment>local-ont-2
</rdfs:comment>
</daml:Ontology>
<daml:Class rdf:ID="bass">
<rdfs:subClassOf rdf:resource="#fish"/>
</daml:Class>
<daml:Class rdf:ID="sub-bass">
<subClassOf rdf:resource="#bass"/>
</daml:Class>
<daml:Class rdf:about="#sub-bass">
<livesIn rdf:resource="#river"/>
<hasHobbyOf rdf:resource="#swimming"/>
</daml:Class>
</rdf:RDF>
```

The above situation has not yet been solved because the strength of the relationship between "*fish*" and "*swimming*" is rather high. In order to solve such problems, WSD should be applied

to the property names as well as the property values: i.e., "*hobby.*"

4   Conclusion

In this paper, we described a perspective for the mapping between the global and the local ontologies based on (Calvanese et al., 2001). We support the global-as-view approach for maintaining the global ontology. This approach is thought to be promising, but it has some unsolved problems concerning inconsistencies in ontologies. Word sense disambiguation, called WSD, can provide one clue for solving these problems because it draws certain close relationships among words appearing as property values or concept definitions. In contrast to (Calvanese et al., 2001), we proposed two situations in which there is ontology inconsistency because one word has multiple meanings: the first case was that of a query based on the wrong choice of meaning for a word, and the second situation involved the inability to use WSD to decide the meaning of a concept. To solve these problems, more precise disambiguation methods should be developed in WSD, or more elaborate representations should be provided for the constraints in ontology definitions. On the other hand, the ontology screening system proposed in this paper can be easily implemented using query engines (e.g., (RQL, 1994)(RDFQL, 2000)(Miller, 2001)(JENA, 2001) based on RDF Query (RDFQuery, 1998) as one of the alternatives. A query language must play a crucial role in the formation of concepts and the screening for inconsistencies over intended ontologies. In the semantic web world, we believe that screening for inconsistencies in ontologies will become critical and that the use of NLP methods like WSD to screen for inconsistencies will offer a significant contribution.

References

T. Berners-Lee, Hendler J., and Lassila O. 2001. The semantic web. *Scientific American*, 279:35–43, May.

D. Calvanese, Giacomo G. D., and Lenzerini M. 2001. A Framework for Ontology Integration. In *The First Semantic Web Working Symposium (SWWS01)*.

DAML+OIL, 2001. *DAML+OIL language.* http://www.daml.org/2001/03/daml+oil.

J. Euzenat. 2001. An Infrastructure for Formally Ensuring Interoperability in a Heterogeneous Semantic Web. In *The First Semantic Web Working Symposium (SWWS01)*, pages 345–360.

JENA, 2001. *The jena semantic web toolkit.* Hewlett-Packard Company, http://www.hpl.hp.com/semweb/jena-top.html.

LDOCE. 1995. *Longman Dictionary of Contemporary English.* Longman.

C. D. Manning and H. Schutze, 1999. *Foundations of statistical natural language processing*, chapter Word Sense Disambiguation, pages 229–261. MIT Press.

L. Miller, 2001. *RDF query using SquishQL.* http://swordfish.rdfweb.org/rdfquery/.

P. Mitra, G. Wiederhold, and Decker S. 2001. A Scalable Framework for the Interoperation of Information Sources. In *The First Semantic Web Working Symposium (SWWS01)*, pages 317–329.

RDFQL, 2000. *RDF Query Analyzer - RDFQL.* Intellidimension, http://www.intellidimension.com/RDF-Gateway/beta3/.

RDFQuery, 1998. *RDF Query Specification.* The World Wide Web Consortium, http://www.w3.org/TandS/QL/QL98/pp/rdfquery.html.

RQL, 1994. *The RDF Query Language.* Institute of Computer Science, Foundation of Research Technology Hellas, http://139.91.183.30:9090/RDF/RQL/.