

Web オントロジのメンテナンスにおける 語の多義性解消からのアプローチ

中挾知延子[†]

本稿では Web 上でのオントロジの一貫性を保つために、自然言語処理技術の一つである多義性解消 (WSD) の適用が有効であることを述べる。オントロジに関しては、データベース分野の視点から、目的のオントロジに対するクエリが重要な役割を果たすと論じられている。理由として、オントロジの中に存在するインスタンスの集合は、複数の概念を満たすインスタンスの集合を求めるクエリの形で表現されるということがある。Calvanese の論文によれば、オントロジをグローバルオントロジと複数のローカルオントロジで構成されるとした場合、グローバルオントロジに存在するデータは、ローカルオントロジに対するクエリの結果、つまりビューとしてとらえることができる。グローバルオントロジとローカルオントロジ間のインスタンスのマッピングを考えたとき、グローバルオントロジのインスタンスはローカルオントロジにある複数の概念を満たすインスタンスと対応付けられる。ただし、その際に正しいマッピングが行われず、結果としてグローバルオントロジにある間違っただけのインスタンスが対応づけられてしまう可能性がある。本稿ではそのような可能性によるオントロジの一貫性の損失を例をあげて説明し、間違っただけのマッピングを防ぐために WSD が適用されるケースについて述べる。

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

Chieko NAKABASAMI[†]

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, 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, and propose that WSD can be one of the promising method to screen such as inconsistencies.

[†] 東洋大学 Toyo University

1. はじめに

近年、新しい Web テクノロジーであるセマンティックウェブ[1]が多くの研究者の注目を集めている[2][3]。データベース分野からのアプローチもさかんに行われており、Web 上の知識を巨大なデータベースとしてとらえ、そのメタデータにおけるオントロジの共有、再利用とこれらのメンテナンスのための研究が進められており[4]、とりわけメンテナンスにおけるデータの一貫性の保持が重要なテーマとなっている。

そのための有効な手法として自然言語処理技術の適用が考えられる。本論文では、Calvanese が論文[4]で主張するグローバルセントリックアプローチにおける、グローバルオントロジとローカルオントロジに関するマッピングの考えに基づいてさえ、メンテナンスにおけるオントロジの一貫性が損なわれるケースが存在することを例示し、その解決策として、自然言語処理における多義性解消の手法が適用できることを述べる[5]。

本稿ではデータベース分野からの研究成果をもとにオントロジを分類し、オントロジを 1 つのグローバルオントロジと複数のローカルオントロジで構成されるとする。ローカルオントロジにはいくつかの概念とそれらの外延であるインスタンスが存在する。一方でグローバルオントロジを、ローカルオントロジにある概念が複数組み合わせられて形作られるビューとしてとらえ、ローカルオントロジとグローバルオントロジのインスタンス間にはマッピングが存在するとしている。

ここで問題となるのは、ローカルオントロジは複数存在し、1 つのローカルオントロジから作られたビューであるはずのグローバルオントロジのインスタンスが、実際のところは複数のローカルオントロジにある概念が組み合わせられたビューとしてグローバルオントロジで認識されてしまうことである。

それらの原因として、概念の名前が複数のローカルオントロジにおいて同じであるために、グローバルオントロジでどの概念を選ぶべきか判断を誤ってしまうことが考えられる。その結果として、グローバルオントロジとローカルオントロジでのインスタンスマッピングが正しいものではなく、ビューで表された概念をローカルオントロジに追加して、既存の概念を拡張したとき、オントロジの一貫性が損なわれてしまう可能性がある。

本稿では、一貫性が損なわれる状況を例をあげて説明し、オントロジの記述方法として DAML+OIL[6]を用いる。

以降の章では、[4]の研究の概略を述べた後、例をあげてグローバルオントロジ、ローカルオントロジ間のマッピングについて説明する。そのマッピングにおいて、オントロジの一貫性が損なわれてしまうケースを述べ、そこに多義性解消を用いることで一貫性が保持できる可能性について論じる。

2. グローバルセントリックアプローチ

グローバルセントリックアプローチ、またはグローバルアズビュー (global-as-view) アプローチは[4]で述べられており、データベース分野でのデータの統合における研究論文[7]で言及されているアプローチである。オントロジをグローバルとローカルなオントロジで構成するとした場合、

1 つのグローバルオントロジと複数のローカルオントロジが存在し、グローバルオントロジとローカルオントロジにある概念間にはマッピングが存在する。グローバルセントリックアプローチにおいて、グローバルオントロジにある概念は、ローカルオントロジにある概念を組み合わせたクエリの結果、すなわちビューとしてとらえられる。このアプローチに対して、ローカルセントリックアプローチがあり、ローカルオントロジにある概念は、グローバルオントロジにある概念を組み合わせたビューとして定義される。たとえば、グローバルオントロジとして図 1 のスキーマを考える。スキーマに従って、図 2 に示す関係がグローバルオントロジとローカルオントロジの間に成り立っているとす。ここで S_1 から S_7 はローカルオントロジにある概念である。図 2 から、ローカルオントロジにある概念 S_1 から S_7 の組み合わせであるビューを満たす概念がグローバルオントロジにおける個々の概念にマッピングされている。[4]では、図 2 を例にとってグローバルオントロジの一貫性が損なわれる 2 つのケースがあげられており、それらを解決する方法や考え方を示している。2 つのケースとは、(1)グローバルなデータモデルが存在しない、(2)複数のデータモデルが存在する、ことである。本稿では、これら 2 つのケースとは違う視点からオントロジの一貫性が損なわれるケースをあげて、それを解く糸口として多義性解消 (Word Sense Disambiguation; WSD) が有効であることを示す。

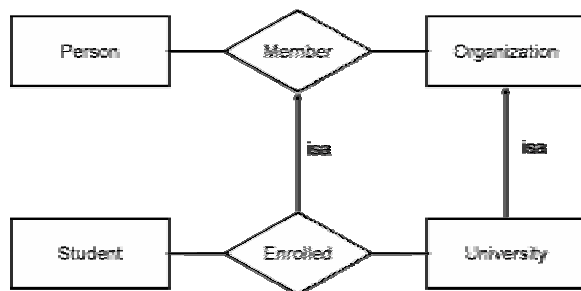


図 1 グローバルオントロジのスキーマ例

Person(X)	$S_1(X)$	
Organization(X)	$S_2(X)$	
Member(X)	$S_6(X,Z)$	$S_7(Z,Y)$
Student(X)	$S_3(X,Y)$	$S_4(X)$
University(X)	$S_5(X)$	
Enrolled(X,Y)	$S_4(X,Y)$	

図 2 グローバルとローカルオントロジとのマッピング例

3. グローバルとローカルオントロジとのマッピング

本章ではオントロジの一貫性が損なわれる 2 つの場合を示し、具体例をあげて説明する。例として図 3 に示すようなグローバルオントロジとローカルオントロジを考える。ここで、グローバルオントロジとローカルオントロジとの概念間には 1 対 1 のマッピングが存在し、それら概念を満たすインスタンスの集合は一致しているものとする。

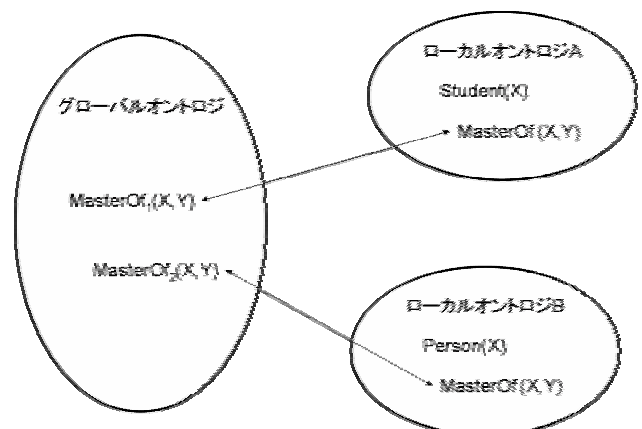


図 3 グローバルとローカルオントロジの例

(1) グローバルとローカルオントロジとのインスタンスのマッピングにおいて、対応するグローバルオントロジのインスタンスが存在しない。

ローカルオントロジ A には Student(X)と MasterOf(X,Y)という 2 つの概念が存在する。ここで概念の意味として、「X は学生である」「X は Y 大学院の修士課程にいる」とする。一方、ローカルオントロジ B には Person(X)と MasterOf(X,Y)という 2 つの概念が存在し、意味としてそれぞれ「X は人である」「X は Y の名手 (ある技能に秀でている) である」とする。このとき、ローカルオントロジ A と B にある MasterOf(X,Y)は項の名前は同じであるが、その中に含まれている意味は異なる。そこでローカルオントロジ A において Student(X), MasterOf(X,Y)という連言は、“X は Y 大学院の修士課程の学生である”ということの意味し、B における Person(X), MasterOf(X,Y)は、“X は Y の名手である人物である”ということの意味する。

ローカルオントロジ A の MasterOf を MasterOf₁, B のそれを MasterOf₂としたとき、それぞれに対応する概念がグローバルオントロジに存在しているとする。ここでローカルオントロジ A からグローバルオントロジに対して以下のクエリが発行されたとする。

Student(X), MasterOf(X, Y) (a)

(a)はローカルオントロジにある概念のビューとしてとらえることができ、(a)のクエリによってグローバルオントロジにある Student(X)と MasterOf₁(X, Y)を同時に満たす、すなわち“修士課程の学生”の集合が返される。ところが、このとき Student(X)と MasterOf₂(X, Y)を同時に満たすインスタンス、つまり“名手である学生”の集合が返された場合、ローカルオントロジ A では一貫性が損なわれるケースが生じる。考えられるケースとして、ローカルオントロジ A には Student のプロパティとして“ある大学に入学している”という Enrolled プロパティが付随し、Enrolled(X, Y)として X は Student のインスタンス、Y は University のインスタンスで構成されているとする。(a)のクエリで返されたインスタンスを Student の下位概念 (たとえば Sub-Student) として分類した場合、“ある大学にいる学生の名手”という意味を持つことになり、グローバルオントロジに再度(a)のクエリを発したとき、もはやグローバルオントロジにはマッピングできるインスタンスが存在しない場合がある。結果としてローカルオントロジ A とグローバルオントロジのインスタンス間の 1 対 1 のマッピングは損なわれる。

(2) グローバルオントロジのインスタンスにマッピングされるローカルオントロジのインスタンスが不完全である。

ローカルオントロジ B から(b)のクエリが発行されたとする。

Person(X), MasterOf(X, Y) (b)

(b)のクエリを発行したとき、グローバルオントロジでは MasterOf₂ が適用されて、“名手である人物 X”を満たすインスタンスの集合が返されるべきである。それに反して、MasterOf₁ が適用さ

れると，“修士課程に在籍する人”を満たすインスタンスの集合が返され，MasterOf₁ プロパティに制約として Enrolled(X, Y)があるため，その制約も継承することになる．

結果として(b)を満たすインスタンスには“大学に入学している”という制約が課され，ローカルオントロジ B とグローバルオントロジにおけるマッピングが不完全なものになる．つまり，“どこかの大学に属していない”名手以外は除外されることになる．

4. WSD の適用

3章で説明したオントロジの一貫性が損なわれる 2 つのケースが，自然言語処理技術の 1 つである WSD の適用により防ぐ可能性を述べる．ローカルオントロジ A と B それぞれが DAML+OIL を用いて図 4，図 5 のように記述されているとする．なお，紙面の都合上，本論文で述べたい内容と直接関係のない箇所，たとえばネームスペースの記述などは省いてある．また，DAML+OIL で先頭に # 記号が付いているクラスオブジェクトの定義に関しては，暗黙の内に記述の別の部分で定義されているものとする．

```
<rdf:RDF>
<daml:Ontology rdf:about="">
<rdfs:comment>localOntology_A</rdfs:comment>
</daml:Ontology>
<daml:Class rdf:ID="Student">
<rdfs:subClassOf rdf:resource="#Person"/>
<rdfs:subClassOf>
<daml:Restriction daml:maxCardinalityQ="1">
<daml:onProperty rdf:resource="#Enrolled"/>
<daml:hasClassQ rdf:resource="#University"/>
</daml:Restriction>
</rdfs:subClassOf>
</daml:Class>
<daml:ObjectProperty rdf:ID="Enrolled">
<rdfs:range rdf:resource="#University"/>
</daml:ObjectProperty>
<daml:ObjectProperty rdf:ID="MasterOf">
<rdfs:comment>Master is a holder of the second
university degree </rdfs:comment>
<rdfs:range rdf:resource="#University"/>
</daml:ObjectProperty>
</rdf:RDF>
```

図 4 ローカルオントロジ A の記述の一部

```
<rdf:RDF>
<daml:Ontology rdf:about="">
<rdfs:comment>localOntology_B</rdfs:comment>
</daml:Ontology>
<daml:Class rdf:ID="Person">
<rdfs:subClassOf rdf:resource="#Animal"/>
</daml:Class>
<daml:ObjectProperty rdf:ID="MasterOf">
<rdfs:comment>Master is a worker who is skilled in
a particular technique and able to teach others
</rdfs:comment>
<rdfs:range rdf:resource="#Technique"/>
</daml:ObjectProperty>
</rdf:RDF>
```

図 5 ローカルオントロジ B の記述の一部

(1) 対応するグローバルオントロジのインスタンスが存在しない場合

3章の(1)のケースを DAML+OIL で表したものを図 6 に示す．図 6 では Student のサブクラスとして Sub-Student が追加され，そのオブジェクトプロパティとして MasterOf が記述され，“名手である学生”が表される．図 6 で追加された Sub-Student クラスにマッピングされるグローバルオントロジのインスタンスを Sub-Student(X)というクエリを発行して取り出した場合，グローバ

ルオントロジにはマッピング先のインスタンス，すなわち“名手である学生”は存在しない可能性がある．このような場合，WSDにより Student と MasterOf プロパティのコメント部分における語との親和性によって判断することが考えられる．

```
<rdf:RDF>
<daml:Ontology rdf:about="">
<rdfs:comment>localOntology_A</rdfs:comment>
</daml:Ontology>
<daml:Class rdf:ID="Student">
<rdfs:subClassOf rdf:resource="#Person"/>...</daml:Class>
<daml:ObjectProperty rdf:ID="Enrolled">
<rdfs:range rdf:resource="#University"/>
</daml:ObjectProperty>
<daml:ObjectProperty rdf:ID="MasterOf">
<rdfs:comment>Master is a holder of the second university degree
</rdfs:comment>
<rdfs:range rdf:resource="#University"/>
</daml:ObjectProperty>
<daml:Class rdf:ID="Sub-Student">
<rdfs:subClassOf rdf:resource="#Student"/>
<rdfs:subClassOf>
<daml:Restriction daml:maxCardinalityQ="1">
<daml:onProperty rdf:resource="#MaterOf"/>
<daml:hasClassQ rdf:resource="#Technique"/>
</daml:Restriction>
</rdfs:subClassOf>
</daml:Class>
</rdf:RDF>
```

図 6 Sub-Student クラスが追加されたローカルオントロジ A

(2) グローバルオントロジのインスタンスにマッピングされるローカルオントロジのインスタンスが不完全である場合

3章における(b)のクエリの結果，図7に示すように Person のサブクラスとして MasterPerson が追加される．MasterPerson クラスには，クラスが持つ MasterOf プロパティの制約として University クラスのオブジェクトを引数にとる制約である“大学の修士課程の学生”が課される．この場合，ローカルオントロジ B にすでに存在する MasterOf プロパティの Technique クラスをさらに狭めた形で University クラスが位置づけられることになる．本来意図した MasterPerson の“特定の技能に秀でた人”という概念にさらに“修士課程の学生”という概念が合わさって，“修士の学生でない特定の技能に秀でた人”は記述されなくなる．ここではWSDによる Technique と University との親和性の判断が必要である．

```

<rdf:RDF>
<daml:Ontology rdf:about="">
<rdfs:comment>localOntology_B</rdfs:comment>
</daml:Ontology>
<daml:Class rdf:ID="Person">...</daml:Class>
<daml:ObjectProperty rdf:ID="MasterOf">
<rdfs:comment>Master is a worker who is skilled in a particular technique
and able to teach others </rdfs:comment>
<rdfs:range rdf:resource="#Technique"/>
</daml:ObjectProperty>
<daml:Class rdf:ID="MasterPerson">
<rdfs:subClassOf>
<daml:Restriction>
<daml:onProperty rdf:resource="#MasterOf"/>
<daml:toClass rdf:resource="#University"/>
</daml:Restriction>
</rdfs:subClassOf>
</daml:Class>
</rdf:RDF>

```

図 7 MasterPerson クラスが追加されたローカルオントロジ B

5. まとめ

本稿では、オントロジを 1 つのグローバルオントロジと複数のローカルオントロジで構成されているとして、グローバルオントロジとローカルオントロジそれぞれに存在するインスタンス間のマッピングに沿ってオントロジの一貫性が損なわれる場合を示した。グローバルオントロジにある概念は、ローカルオントロジにある概念のビューであるとしたグローバルセントリックアプローチに従って、クエリの発行によってビュー概念を満たすインスタンスをグローバルオントロジから 1 対 1 のマッピングによって得られるとした。その際、異なるローカルオントロジに存在する同じ名前の概念によって生じるオントロジの一貫性の損失について 2 つのケースを示し、DAML+OIL を用いて例示した。クエリの結果がサブクラスとしてローカルオントロジに追加されるときに起きる問題は WSD による語同士の親和性により解決できることを示し、具体例をあげて説明した。

WSD による判断の際には、オントロジ記述におけるコメント部分など、自然言語で記述した部分が重要な意味を持つと思われる。言い換えれば、豊かな表現力を持つ自然言語で記述できる部分をオントロジに多く加えることによって、WSD のみならずさまざまな自然言語処理技術を駆使してオントロジのメンテナンスが改善される。セマンティックウェブでのメタデータの表現としての RDF[8]に、より多くの注釈を自然言語で付けることで、人に優しい質問形式で Web から検索結果を得られるようにした研究[9]も注目すべきである。今後の研究として、オントロジの一貫性を保持するための WSD をはじめとした自然言語処理技術を実証することと、RuleML[10]のようなオントロジにおけるルール記述と推論機構も取り入れて、オントロジメンテナンスのためのシステムの開発を予定している。

参考文献および参照 URL

- [1] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*. 279:35–43, May, 2001
- [2] 情報処理学会編 . 特集セマンティック Web . 情報処理 . Vol.43, No.7, pp.707-750. 2002.7
- [3] 人工知能学会編 . 特集 : 「 Semantic Web とその周辺 」 . 人工知能学会誌 . Vol. 17, No.4, pp.383-416. 2002.7
- [4] D. Calvanese, Giacomo G. D., and Lenzerini M. A Framework for Ontology Integration. *Proc. of the 1st Semantic Web Working Symposium(SWWS01)*. pp.303-316. 2001
- [5] Chieko Nakabasami and Naoyuki Nomura. A Proposal for Screening Inconsistencies in Ontologies based on Query Languages using WSD. *Proc. of the 2nd Workshop on NLP and XML (NLPXML-2002)*. pp.109-113. 2002
- [6] DAML.org. Index of /2001/03. <http://www.daml.org/2001/03/>
- [7] R. Hull. Managing Semantic Heterogeneity in Databases: A Theoretical Perspective. *Proc. of the 16th ACM SIGACT SIGMOD SIGART Symposium on Principles of Database Systems (PODS'97)*. pp.51-61. 1997
- [8] Resource Description Framework (RDF) / W3C Semantic Web Activity. <http://www.w3.org/RDF/>
- [9] B. Katz and J. Lin. Anotating the Semantic Web Using Natural Language. *Proc. of the 2nd Workshop on NLP and XML (NLPXML-2002)*. pp.37-44. 2002
- [10] RuleML Home page. <http://www.dfki.uni-kl.de/ruleml/>