

オントロジ構築による異種スキーマ間の検索

丸山朝也 塚本享治[†]

東京工科大学 大学院 バイオ・情報メディア研究科 メディアサイエンス専攻[‡]

1. はじめに

同じような内容を持つ XML 文章であっても、スキーマが異なると共通したクエリにより検索することができない。そのため、共通したクエリにより異なるスキーマを持つ XML 文章を検索するためには、各スキーマから切り離した共通の構造を定義し、その構造と各スキーマとの対応を関係付けることにより、発行されたクエリを各スキーマにマッチしたクエリへと変換する方法が考えられる。

本研究では、大学における複数学部のシラバスを題材とし、それぞれのシラバスに共通する構造のオントロジを構築した。また、そのオントロジと、各シラバスとの対応関係を記述することにより、オントロジに対するクエリを各シラバスに対するクエリへと変換・検索・抽出できるシステムを構築した。

2. オントロジの構築

2.1. オントロジのレベル

オントロジとは、各ドメインに対する表現や構造などの概念を体系化したものであり、記述されたオントロジを理解することで、対象ドメインの概要を体系的に理解できるものである。オントロジには、対象の知識構造を表す内容的な部分と、META 情報などを表す構造的な部分との二つのレベルに大きく分けることができる。

内容的なオントロジを構築することにより、表現の異なる表現対象を、同一な概念であると機械的に把握できたり、概念間の関係を機械的に把握できたりする。例えば、コンピュータと計算機は、ほぼ同じ概念を表していることや、ある科目を履修する場合、別の科目を以前に履修している必要があるといった依存関係などを

機械的に把握することができる。

構造的なオントロジを構築することにより、各スキーマ内の要素が、どのような概念を表す要素なのかを把握することができる。XML 文章のように階層化されたデータである場合、要素が出現する位置によって、同一の要素名が使われていたとしても、同じ概念を表す要素であるとは限らない。例えば、XML 文章で、「名前」という要素名が二つ異なる位置で出現したとする。一つは、「講師」要素の下位の要素として出現し、もう一つは、「科目」要素の下位の要素として出現したとする。この場合、前者は講師名を表す要素となり、後者は科目名を表す要素となる。このように、構造を表すオントロジを構築し、スキーマ内の要素とオントロジを関係付けることにより、何を表す要素なのかを機械的に把握することができる。

2.4. オントロジの利用

本研究では、大学におけるシラバスを題材とし、構造を表すオントロジをオントロジ言語である RDF[1], RDFS[2], OWL[3]を用いて構築した。

図1は、内容を表したオントロジ・構造を表したオントロジ、リソース間の関係を表したものである。

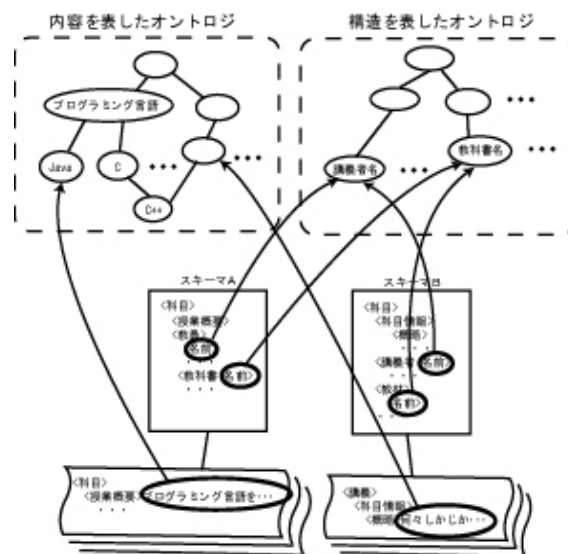


図1 オントロジとリソース間の関係

Retrieval of different schemata by construction of ontology

[†]Asaya Maruyama Michiharu Tsukamoto

[‡]Tokyo University of Technology

Graduate School of Bionics, Computer and Media Sciences
Media Science Program

内容的なオントロジと、構造的なオントロジでは、対応関係を定義すべき対象が異なる。内容的なオントロジでは、抽出された情報とオントロジの対応する概念を見るか、オントロジを見て、対応する情報を抽出するかである。それに対し、構造的なオントロジでは、スキーマ内にある要素の位置情報とオントロジを関係付けることにより、要素の意味をオントロジから知るか、オントロジから要素を特定することが可能となる。本研究では、対象となる情報を XML 文章と限定し、位置情報を記述するために、XPath[4]を用いた。

図2は、オントロジと XPath により表現されたスキーマ内にある要素の位置情報を結び付けている例である。ここでは、スキーマ内の「/科目/概要」に位置する要素は、構造的なオントロジ内の「科目概要」という概念と等しいと定義している。

```
<owl:Class rdf:about="syllabus:/科目/概要">
  <owl:equivalentClass rdf:resource="common:科目概要"/>
</owl:Class>
```

図2 XPathを用いた位置情報の定義例

本研究では、このようなオントロジと位置情報を対応できる表を、各学部のシラバスごと（スキーマごと）に RDF, OWL を用いて定義した。また、定義する際に必要最低限の記述量ですむように、RDF 内で変数を利用できるようにした。

図3は、RDF 内で変数を定義し、利用している例である。この例では、“executeNumber”という変数を宣言し、値として“第1回目”をいれる。次に、XPath のガード内で先ほど定義した変数を利用している。

```
<!-- 変数の定義と初期化 -->
<owl:Class rdf:about="variable:${executeNumber}=第1回目">
  <owl:equivalentClass rdf:resource="common:実施第1回目"/>
</owl:Class>

<!-- 変数の利用 -->
<owl:Class rdf:about="syllabus:/科目/授業[@回数=${executeNumber}]">
  <owl:equivalentClass rdf:resource="common:授業"/>
</owl:Class>
```

図3 RDF内での変数定義と利用の例

3. システムの構築

本研究では、構造的なオントロジ内の概念を選択することにより、登録されている XML 文章から、スキーマ別に位置情報を特定し、検索するシステムを構築した。

システムが行う処理の流れを簡単にではあるが、以下に示す。

- 1) オントロジ内の概念で抽出したいものを選択し、システムに通知する。
- 2) システムは、オントロジとの関係を定義した RDF から該当するものを検索する。この検索には、Jena フレームワーク[5]を用いた。
- 3) もし変数を定義しているようであれば、システム内に変数を登録する。
- 4) XPath を取得する。変数が利用されている場合は、システム内に登録されている変数値と置き換える。
- 5) 取得した XPath を用いて XML 文章を検索する

4. 評価実験

評価のため、シラバスの構造的なオントロジを構築した。また、構築したオントロジとの関係が定義された RDF を東京工科大学にあるシラバスのスキーマの数だけ用意した。これらを用い、構築したシステムで東京工科大学のメディア学部、コンピュータサイエンス学部、バイオニクス学部のシラバスを検索したところ、オントロジへのクエリを、各スキーマに対応したクエリへと変換し、3 学部のシラバスからの検索を正常に行えることが確認できた。

5. おわりに

本研究では、構造的なオントロジを構築し、そのオントロジとスキーマを関係付けることにより、異なるスキーマを持つ XML 文章であっても同一のクエリにより検索できることが確認できた。

今後、今回確認できた機構を発展させ、スキーマが異なるが、内容的には同じような XML 文章間での変換ロジックの自動生成を行いたいと考えている。

- [1] W3C, “Resource Description Framework”, <http://www.w3.org/RDF/>, 2005
- [2] W3C, “RDF Vocabulary Description Language 1.0 RDF Schema”, <http://www.w3.org/TR/rdf-schema/>, 2004
- [3] W3C, “OWL WEB ONTOLOGY LANGUAGE”, <http://www.w3.org/TR/owl-features/>, 2004
- [4] W3C, “XML Path Language (XPath)”, <http://www.w3.org/TR/xpath/>, 1999
- [5] Hewlett-Packard Development Company, “Jena - A Semantic Web Framework for Java”, <http://jena.sourceforge.net/>, 2005