

オントロジーマッピングによる LOM と CRM 間のメタデータ変換

立川 正和^{*1} 内田 邦晃^{*2} 澤井 進^{*3} 若木 利子^{*2}

芝浦工業大学大学院 工学研究科^{*1} 芝浦工業大学 システム工学部^{*2} (財)学習ソフトウェア情報研究センター^{*3}

1 はじめに

現在、国立科学博物館や NICER 等において標本データ又は教材のアーカイブを蓄積し、それらに付与した CRM 準拠の標本メタデータに基づく S-Net[7] や LOM に基づく NICER 検索システム [6] が稼動している。両者において互いのアーカイブを利用したいというニーズがあるが、両者の概念が異なることに起因する単純な対応が見つからない、さらにメタデータに潜む情報欠落に対する知識の補完といった技術的課題に対処する必要がある。ところで、Web 上にメタデータやオントロジーを表現するための技術であるセマンティック Web の RDF 及び OWL[3] が W3C により勧告された。他方、異なるオントロジーを対応させ変換を行うオントロジーマッピングが近年試みられている。本研究では、各メタデータをセマンティック Web の RDF で表現したオントロジーを用いて両者の概念を対応させると共に独自に OWL で表現した生物分類オントロジーを構築、それを用いて知識の補完に基づくオントロジーマッピングを行い相互変換を実現し有用性の評価を行った。

2 LOM と CRM 間のオントロジーマッピング

2.1 LOM と CRM の RDF 表現

本研究では LOM と CRM 準拠の 2 種類のメタデータを対象とする。LOM(Learning Object Metadata)とは IEEE により 2002 年に認証された学習オブジェクトに関するメタデータの構造と語彙指定の定義する規格であり、96 項目が定められている。この RDF 表現については [1, 4] で提案した。CIDOC CRM(Conceptual Reference Model)[2] とは国際博物館会議で定められた博物館の収蔵品管理に関するオブジェクトの情報を表すモデルであり、OWL 表現で提案された。国立科学博物館では本規格に準じた 70 項目からなるメタデータが定められている [7]。

表 1: LOM(上) と CRM(下)

タイトル	URL	概要	...
トラフグ	http://~	この魚~	...
DB URL	和名	界名	...
http://~	トラフグ	Animalia...	...

Metadata translation between LOM and CRM based on ontology mapping

^{*1}M. Tachikawa, ^{*2}K. Uchida, ^{*3}S. Sawai, ^{*2}T. Wakaki

^{*1}Graduate School of Engineering, Shibaura Institute of Technology

^{*2}College of Systems Engineering, Shibaura Institute of Technology

^{*3}Software Information research center for Learning

2.2 オントロジーマッピング

LOM と CRM は異なる概念に基づくメタデータである為、そのままでは互換性を持たない。そこで両者をオントロジーで表し、対応付け変換を行うオントロジーマッピングを行った。対応付けは 1 つの項目に対し 1 つが対応するとは限らず以下の 3 つに場合分けされる。

1. LOM の 1 項目に対し CRM の 1 項目に対応させる。
2. LOM の多項目に対し CRM の 1 項目に対応させる。
3. LOM の 1 項目に対し CRM の多項目に対応させる。

本研究では主に 1 の場合と 3 の場合が多く割合を占め、3 の代表例としては LOM のキーワードに対する CRM の学名、和名、採集に関する項目がある。オントロジーマッピングのイメージを図 1 に示す。

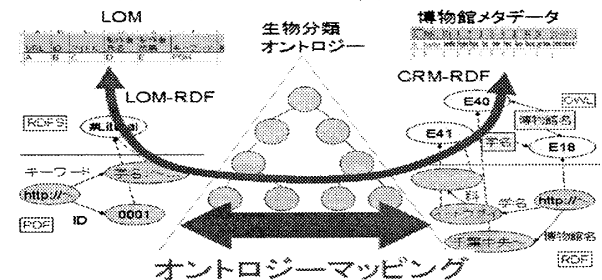


図 1: オントロジーマッピングのイメージ

2.3 生物分類オントロジーを用いた欠落した知識の補完

オントロジーマッピングを行う場合、項目を対応させるだけではなく値から対応する項目を推定する必要となるものもあり、正しくマッピングを行わないと情報欠落が発生してしまう。本研究においては生物分類に関する項目において値が欠ける、対応する項目がはっきりしないなど何らかの形で欠落した知識の補完を行う必要がある。生物の分類とは生物を階層構造により体系的に分類するもので、上位階層から界、門、綱、目、科、属、種の順に分類する。例えばトラフグなら動物界 脊索動物門 硬骨魚綱 フグ目 フグ科 トラフグ属 トラフグ種に分類される。本研究ではこの階層構造に着目し、その構造を利用して欠落した知識の補完をするためのオントロジーを OWL 言語を用いて表現し構築した。

- 生物の分類として界から種のどれに存在するか、また分類名が学名か和名かを表す情報を OWL クラスとして表す。各分類名は分類と言語の情報を持ったため、分類と言語を表す OWL クラスの積をとって新しくクラスを表現する (図 2(1))。
- 各分類名は積をとったクラスに属する個体として表現する。例えばフグ科は分類が科かつ言語が和名であるのでこの積をとったクラスを表現し、それに属する個体とする (図 2(2))。

- 分類の上下関係は superClassification プロパティを用いて関係付け、この上下関係は包含関係であるのでプロパティに推移性を持たせる (図 2(3)).
 - japaneseName, scientificName(逆プロパティの関係)を用いて学名と和名を関係付ける (図 2(4)).
 - 和名は違う階層での名前重複が発生するため和名に番号を付けて重複を回避し、word プロパティを用いて共通のキーワードを関係付ける (図 2(5)).
- 補完の手順は以下のようにする。

1. LOM-RDF 又は CRM-RDF から既知の値の中で一番下位にあたる学名又は和名を取得する。和名を取得した場合はそれに対応する学名を生物分類オントロジーを用いて取得する。
2. 取得した学名の1つ上位の分類に対して値 (インスタンス)があるかどうか調べ、値が存在する場合は知識の補完を行わずに、被対象となる値を調べた値に変更する。値が存在しない場合は、生物分類オントロジーを用いて補完の被対象となる学名の上位の分類を検索、取得し値を埋める。
3. 一番上位の階層の分類に対して処理を終えるまで1つずつ上位の分類を調べ知識の補完を行う。

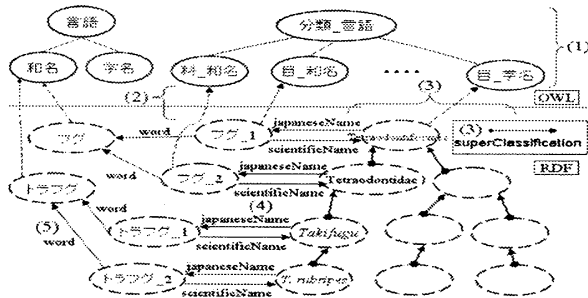


図 2: 生物分類オントロジーのイメージ

3 実装と評価

LOM と CRM 間のマッピングには Java を用いて変換ツールの実装を行った。変換時における RDF の処理や OWL で構築した生物分類オントロジーを用いた推論や知識の補完時には OWL の推論を行うことができる Jena (Java API)[5] を用いた。

評価として、まず、三重大学水産実験所の CRM データ 1000 件を用いて知識の補完に関する評価を行った。表 2 に生物分類オントロジーに関する学名の項目について補完前後の埋まっている件数を示す。界、門、綱、目についてデータが皆無であったものが生物分類オントロジーを用いることで、オントロジーを構築していない 1 件を除き知識の補完ができていくことがわかる。

表 2: 三重大学のデータでの補完前と補完後の比較

	界	門	綱	目	科	属	種
補完前	0	0	0	0	1000	999	999
補完後	999	999	999	999	1000	999	999

次に、知識の補完に対する評価を行った。図 3 は三重大学の魚に関する標本データであるベニツケギンポと呼ばれる魚から界名～種小名を検索、取得した結果である。補完前は S-net の標本情報検索システム、補完後は RDF を SeRQL と呼ばれる SQL ライクに検索を行う

図 3: 補完前(上)と補完後(下)の検索の比較

ツールである Sesame[8] を用いて検索を行った。S-net ではいくつかの項目に分類に関するデータが登録されていないため取得できていないが、補完した結果全ての分類名を取得できていることが分かる。

最後に LOM から CRM へオントロジーマッピングを行った例を示す。図 4 は *Thunnus alalunga* (和名: ビンナガ) についての LOM と、それを CRM へマッピングを行った例である。この標本は CRM に登録されていないため LOM からマッピングを行えば有用であるが、元の LOM には生物の分類が記述されておらず単純に変換ができない。そこでビンナガという和名から生物分類オントロジーで知識の補完を行うことで分類に関する項目が補完され CRM として扱うことができる。

品名	ビンナガ
概要	胸びれが非常に長く、体は白身でかなり淡白です。毎歳1センチほどに加工されています。体長1m前後、6月頃産卵します。九州近海でよく漁獲されています。夏が旬です。関東地方ではビンナガ、関西地方ではトンボと称されています。
キーワード	シーチキン、ビンナガ、ビンナガ魚
界名	動物
門名(日本)	動物
綱名(日本)	脊椎動物
目名(日本)	硬骨魚
科名	マグロ
属名(日本)	マグロ
種小名(日本)	alalunga
種名(日本)	ビンナガ

図 4: LOM(上) から CRM(下) へのマッピングした例

4 おわりに

構築した生物分類オントロジーを用いて LOM と CRM 間のオントロジーマッピングを行い、NICER の教材コンテンツと博物館の標本コンテンツの相互利用・検索を可能とし、他方、不足したメタデータ値の知識の補完を行うことにより S-Net 等で検索できるコンテンツの増大が期待できる。

謝辞

三重大学水産実験所の木村清志教授と国立科学博物館の井上課長、松浦研究員より本研究で使用した CRM データを頂き深く感謝致します。

参考文献

- [1] 立川 正和, 澤井 進, 若木 利子: セマンティック Web 技術を用いた e-Learning 支援システム - LOM-RDF 変換ツールの実装 -, 情報処理学会第 68 回全国大会講演論文集, 5L-5, pp. 221-222, 2006.
- [2] 秋元 良仁: 博物館・美術館の収蔵品管理用メタデータ管理, デジタル図書館ワークショップ第 25 回, 2004.
- [3] G. Antoniou, F.v.Harmelen: A SemanticWeb Primer, The MIT Press, 2004.
- [4] S. Sawai, T. Wakaki, M. Tachikawa, and S. Kunifuji: Multimedia Portals based on Semantic Web Technology for GENES Communities - The Implementation of LOM-RDF Conversion Tools - Proceeding of SWAMM, WWW2006 Workshop, 2006.
- [5] Jena: <http://jena.sourceforge.net/>
- [6] NICER: <http://www.nicer.go.jp/>
- [7] S-net: <http://www.science-net.kahaku.go.jp/>
- [8] Sesame: <http://www.openrdf.org/>