

# 図書館分類法へのオントロジーマッピング手法の適用可能性\*

和田匡路 (学籍番号 200921749)

研究指導教員：緑川信之

副研究指導教員：谷口祥一

## 1. 研究背景・目的

近年、オントロジー工学, SemanticWeb 等の研究領域で、「オントロジーマッピング (Ontology Mapping)」が研究されている [1]。これは、オントロジーのクラス間の対応関係を発見すること、また発見手法の開発を目的としている。対応関係とは、それらが同じであるまたは類似しているという関係である。

オントロジーは様々に定義されているが、それらの中には図書館分類法もオントロジーとみなせるような定義も存在する。このことから、オントロジーマッピング手法は図書館分類法間のマッピングにも適用できると考えられる。そこで本研究では、オントロジーマッピング手法の図書館分類法への適用可能性を検証した。

図書館分類法として日本十進分類法と国立国会図書館分類表を取り上げ、両者の間のマッピングに市瀬らが開発したオントロジーマッピングシステム HICAL で用いられた手法 (以下, HICAL の手法) [2] を適用する実験を行った。

## 2. オントロジーマッピング手法

### 2.1 HICAL の手法

HICAL の手法の基本的なアイデアは、異なるオントロジーの 2 つのクラスが同一のインスタンスを多く含むならば、それらのクラスは似ている、つまり対応関係にあると判定する、というものである。本研究では、オントロジーが図書館分類法、インスタンスが既に図書館分類法で分類された書誌データにあたる。

マッピングは 2 つの図書館分類法の最上位のクラスからその下位のクラスを順次判定してゆくことで行われる。クラス同士が対応関係にあるか判定し、

それらが対応関係にあればその下位のクラスも対応関係にある可能性が存在すると考える。逆に上位のクラス同士が対応関係になれば、その下位のクラス同士も対応関係にないと考える。以降、マッピングの具体的な流れを示す。

まず、2 つの図書館分類法の最上位のクラス同士が対応関係にあるか否かの判定を行う。対応関係にあると判定された場合は、そのペアを記録するとともに、次に対応関係にあるか判定する候補を作成する。対応関係にあると判定したペアの下位クラス間も判定するために、次のように候補を作成する。対応関係にあるクラス A, B に対して作成する候補は次の 3 つによりできる全てのペアである。ただし、「A(B)の子」とは「A(B)の一つ下のクラス」を指す。

- A と B の子のペア
- A の子と B のペア
- A の子と B の子のペア

これらの候補も同様に対応関係にあるか判定し、対応関係にあると判定されたら新たな候補を作成する。作成される全ての候補が判定され、新たに候補が作成されなくなったら、記録していた対応関係にあるペアを出力し、マッピングを終了する。

対応関係にあるか否かの判定は、クラスが含んでいる書誌データの数を元に、k 統計量を用いて行われる。ただし、図書館分類法の階層構造を利用し、あるクラスより下位のクラスに位置づけられたインスタンスは、すべてそのクラスに位置づけられたものとして判定する。これを本研究では「まとめ上げ」と呼ぶ。

### 2.2 改変手法

HICAL の手法では、そのアルゴリズム故に下位クラス間ほどマッピングされない可能性が高いと考えられる。そこで、下位クラス間のマッピングを行う、HICAL の手法の改変手法を考案し、これも適用可

---

\* “Application of Ontology Mapping Techniques to Library Classifications” by Masamichi WADA

能か検証することにした。改変手法は次のようにマッピングを行う。

- I) 2つの図書館分類法の分類記号が記載された書誌データを収集する。
- II) 個々の書誌データで共出現した2つの分類記号の組み合わせを抽出する。
- III) 抽出した分類記号の組み合わせが示すクラス同士のペアを、全て対応関係にあるか否か判定する。判定方法はHICALの手法と同じくk統計量を用いるが、「まとめ上げ」の処理は行わない。

### 3. 実験結果

日本十進分類法と国立国会図書館分類法のマッピングを、J-BISC DVD版(2009)の書誌データを用いて行った。

2つの図書館分類法の全クラス間のマッピングを行うのは困難である。そこで、『図書館情報学ハンドブック第2版』を参考に、NDCの「007(情報科学)」と「010(図書館. 図書館学)」とその下位クラスを、図書館情報学を扱ったクラスであるとみなし、それらを含む書誌データのみを用いた。

#### 3.1 HICALの手法の結果

285ペアが対応関係にあると判定された。この中には、実際に書誌データが分類されるようなNDCの本表と補助表を合成したクラスも含まれている。対応関係にあると判定されたNDLCのクラスは「U(学術一般)」又は「M(科学技術一般)」とその下位クラスのみであった。

#### 3.2 改変手法の結果

87ペアが対応関係にあると判定された。対応関係にあると判定されたNDLCのクラスは「U」「M」以外に、「D(産業)」「F(教育)」の下位クラスも存在した。

85ペア中、HICALの手法による285ペアと一致したペアは69ペアであった。

#### 3.3 手作業によるマッピングとの比較

2つの手法によるマッピング結果がどのようなものかはわからない。一方で、人間が手作業でマッピングをするといったこともこれまでには行われてきた。そこで、NDCの図書館情報学の領域を扱ったクラスを手作業によりNDLCのクラスにマッピングした(181ペア)。これと各手法の結果を比較した(手作

業ではNDCの本表+補助表のクラスはマッピングしていない。比較する結果からもそれらのクラスを含むマッピング結果は除外して比較を行った)。

#### 1) HICALの手法との比較

HICALの手法で対応関係にあると判定された197ペアのうち、3割程度(64ペア)が手作業の結果と一致した。

#### 2) 改変手法との比較

改変手法で対応関係にあると判定された64ペアのうち、6割程度(38ペア)が手作業の結果と一致した。

#### 3) HICALの手法+改変手法との比較

2つの手法の結果をまとめた213ペアのうち、3割程度(70ペア)が手作業の結果と一致した。

## 4. 考察・結論

### 4.1 HICALの手法について

対応関係にある多数のペアを得たことから、手法を適用することはできたと考えられる。ただし、対応関係にあるペアのうち、手作業のマッピング結果と一致したのは3割程度だった。

手作業と一致しなかったペアには、手作業では発見が難しいと思われるもの、手作業のマッピング次第によっては、手作業と一致したと考えられる対応関係が含まれた。このことから、手作業でマッピングするときの支援などとして活用できると考えられる。

### 4.2 改変手法について

対応関係にあるペアは得たものの、その数は少なかった。しかし、HICALの手法とは異なる方法でマッピングを行い、手作業のマッピング結果と一致する割合は高かった。このことから、他の手法によるマッピングの補助として活用できると考えられる。

## 文献

- [1] Euzenat, Jérôme.; Shvaiko, Pavel. *Ontology Matching*. Springer-Verlag, 2007, 333p.
- [2] 市瀬龍太郎ほか. 階層的知識間の調整規則の学習. *人工知能学会論文誌*. 2002, vol. 17, no. 3, p. 230-238.