

オントロジの構築方法に関する研究

2002MT036 川出 裕明 2002MT042 前田 竜臣

指導教員 青山 幹雄

1. はじめに

セマンティック Web の主要技術に語彙や概念間の関係を表現するオントロジがある[5]。オントロジは機械が認識し処理する技術であるので人間にとって表現や理解が困難である。本研究はこの問題点を解決する方法として、UML を利用したオントロジの構築方法を提案する。そして、電子商取引サイトにおけるカテゴリ間の関係をオントロジで表現することを考え、UML を利用したオントロジの構築方法についての考察を行う。

2. オントロジ

2.1. オントロジとは

オントロジとは概念間の関係を表したものである。本研究ではオントロジ記述言語として OWL(Web Ontology Language)を使用する。OWLは推論可能な記述方法が豊富であり、広く普及している。OWL では概念を表すクラス定義とその関係を表すプロパティ定義を行うことができる。また、プロパティの主語をドメイン、目的語をレンジと呼ぶ。

2.2. オントロジ構築上の問題点

オントロジには多様な知識を記述する表現力が要求される。そのため論理的な記述が必要となるが、複雑な構造となってしまう。それが人間には理解が困難となり、オントロジ普及の妨げとなる。

2.3. 解決策

この問題を解決するために本研究ではオントロジ表現を視覚化する。オントロジの論理的構造を UML で表現して人間に理解しやすいオントロジ構築方法を提案する。

3. UML マッピング

3.1. UML を用いる目的

UML とはオブジェクト指向のモデル表記法である[3]。UML をオントロジ構築で利用する理由を以下で示す。

(1) 図的表現

図的表現は人間に文字表現よりも理解しやすく、一目見ただけで全体的な構造を知ることができる。

(2) 開発環境

アプリケーション開発者が UML を利用する場合、オントロジを同じ開発環境で構築できるので、コストを減らせる。

3.2. UML でのオントロジ構築の問題点

オブジェクト指向とオントロジは概念間の関係を表すためにクラス階層を取るという点で相似である。しかしそれぞれの表現方法には相違点もあるため、オントロジを UML で表現する場合オブジェクト指向では扱わない概念を表現しなければならない。従って UML の拡張が必要である。

UML の拡張方法にステレオタイプがある。<<>>で囲まれた所にオントロジで使用される概念を記述することでオントロジの性質を UML で表現できる。

4. OWL の UML 表現

4.1. UML での表現方法

OWL を基本表現、クラス表現、プロパティ表現の3つに分類して順に UML に対応付ける[2]。

4.2. 基本表現

クラスとプロパティの存在を示す。

4.2.1. クラスに関する基本表現

次の4つの公理でオントロジのクラスの存在を示す[1]。

- 1) クラス存在公理
名前を付与してクラスの存在を示す。
- 2) サブクラス関係公理
あるクラスが別のクラスのサブクラスであることを示す。
- 3) 等価関係公理
二つのクラスが等しいことを示す。
- 4) 背反関係
二つのクラスが互いに素の関係であることを示す。

UML ではクラス存在公理をクラスで、サブクラス関係公理を継承関係で表現できる。また、等価関係と背反関係はオブジェクト指向では使用されない概念であるため、ステレオタイプを用いて表す。

4.2.2. プロパティに関する基本表現

プロパティ存在公理で生成されたプロパティが存在することを示す。プロパティはデータ型プロパティとオブジェクトプロパティの2種類に分別できる。データ型プロパティはレンジをデータ値に、オブジェクトプロパティはレンジをクラスに結びつけるプロパティである。

データ型プロパティは主語に対して値の範囲のあるデータ型に制限するので属性に対応付けできる。また、オブジェクトプロパティはクラスとクラスの関係を表すので関連

名に対応付けできる。

4.3. クラス表現

OWLではクラスの関係に加えてクラスの性質を記述して制約を受けた新たなクラスを定義できる。

(1) 値制約を持つクラス

あるクラスに適用されるプロパティの値域に制約を設け、それを満たす個体の集合からなる仮想的クラスである。

オブジェクト指向ではこの概念が存在しないため関連等とステレオタイプを組み合わせて表す。

(2) 出現回数制約をもつクラス

ある特定のプロパティに対する定義個数の制約を課した仮想的クラスである。関連の対象の個数は多重度で表現するため、この制約はステレオタイプと多重度で表現できる。

(3) 集合クラス

集合の論理関係をクラスに適用することで定義されるクラスである。論理関係によって作られたクラスは元となるクラスに依存するため“依存”で表す。そしてステレオタイプで論理組み合わせを記述する。特に和集合によるクラス表現はUMLでは“集約”と対応付けできる。

(4) 列挙クラス

クラスに含まれるインスタンスを過不足なく列挙することで表現されるクラスである。UMLの列挙型の記述方法と同様に属性にインスタンスを記述して対応付けできる。

4.4. プロパティ定義

OWLではプロパティ自身を定義できる。データ型プロパティはクラス、オブジェクトプロパティは関連クラスで定義され次の4つで示される。

(1) 関係公理

プロパティ間の関係を定義する公理である。等価関係公理と逆関係公理がある。ともにクラス間の関連にステレオタイプを用いて表す。

(2) 個数制約公理

プロパティの出現回数をグローバルに制約する公理である。これはプロパティ自身に内在する性質であるのでクラスにステレオタイプを用いて表現する。

(3) 論理的公理

プロパティに論理的性質を持たせることで推論を可能とする公理である。これもプロパティ自身に内在する性質といえるのでクラスにステレオタイプを用いて表現する。なお、この公理はオブジェクトプロパティのみ有効である。

表1と表2でOWLのUML表現の対応関係を示す。

表1 基本表現

クラス公理	クラス定義	クラス
	サブクラス定義	汎化関係
	等価・背反関係	ステレオタイプ
プロパティ公理	データ型プロパティ	属性名
	オブジェクトプロパティ	関連名

表2 クラスとプロパティ表現

クラス表現	値制約	関連, 属性, 型と初期値
	出現回数制約	多重度
	集合	依存, 集約
	列挙	列挙クラス
プロパティ表現	関係公理	関連とステレオタイプ
	個数制約公理	ステレオタイプ
	論理的公理	ステレオタイプ

5. 電子商取引におけるオントロジの構築

5.1. 電子商取引におけるオントロジの利用

Webを利用した電子商取引サイトが数多く存在する。各サイトは商品をカテゴリ化することによって商品情報を整理している。分類方法は各サイトによって異なるため、これらを関連付けて比較するのは難しい[4]。そこで各サイトのカテゴリ階層をオントロジで対応付けて関係を記述し、一つのポータルサイトで管理することを考える。その結果ポータルサイトを通してサイト間のカテゴリを比較できる。しかし、カテゴリ階層間の関係を記述するとき、そのカテゴリが扱う商品の範囲の特定が困難である。この問題を考慮したオントロジの構築方法を提案する。

5.2. カテゴリの範囲の判定方法

カテゴリ間の関係を判定する尺度として、[1]式で示すような一致度を提案する[6]。

$$\text{一致度} = \frac{\text{比較先カテゴリと一致する要素数}}{\text{比較元カテゴリの持つ要素数}} \quad [1]$$

一致度は二つのカテゴリ間の双方向で計算する。ここで扱う要素数はカテゴリを階層全体と見なしたときには下位カテゴリ数を指す。また、一つのカテゴリと見なしたときにはそのカテゴリに含まれる商品数を指す。この値にしきい値を設けてカテゴリ間の関係を調べる。しきい値が高いほど等価関係となる条件が厳しくなる。二つのカテゴリにそれぞれ一致度を適用したとき、その関係は図1になる。

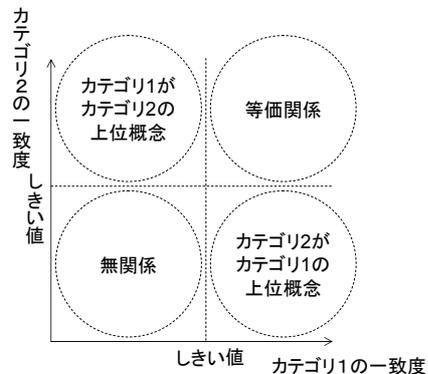


図1 一致度の関係を表した図

5.3. 観点の違いを考慮したクラスの定義

観点の違いを発見するためにはカテゴリに含まれる商品データを見る必要がある。商品名に含まれる単語から各カテゴリの性質を見つけることができる。その単語をレンジとするプロパティを定義し、プロパティ制約を課したクラスを定義することで観点の違いを考慮したクラスを定義できる。

5.4. オントロジの構築手順

以下のような手順でオントロジを構築していく。

(1) 二つのサイトのカテゴリ階層をクラス図で記述

カテゴリマッピングを行うためにはまず二つのサイトのカテゴリ階層の全体像を知る必要がある。そのため、クラス図を使ってそれぞれのサイトのカテゴリ階層を表現する。

(2) 二つのカテゴリ階層の最下位カテゴリをマッピング

下位のカテゴリのほうが扱っている範囲が狭くカテゴリ名から商品の扱う範囲も特定しやすい。従って個人の考え方のずれは少ないと考えられるので、最下位カテゴリ間のマッピングは人間の判断で行う。

(3) 一つのカテゴリを選択し比較先のカテゴリ階層で最も一致度の高いカテゴリとマッピング

まだマッピングされていない一つのカテゴリを選ぶ。これを基準カテゴリとして比較先のサイトのカテゴリ階層で最も類似しているカテゴリを選び二つのカテゴリをマッピングする。カテゴリ間の類似関係の判定に一致度を用いる。

6. 実例を用いたカテゴリマッピング

6.1. カテゴリレベルでのマッピング

以上の構築手順を元に書籍の電子商取引サイトの Amazon と bk1 のカテゴリ階層構造のマッピングを行った。図2と図3に Amazon と bk1 の階層構造を示す[7][8]。

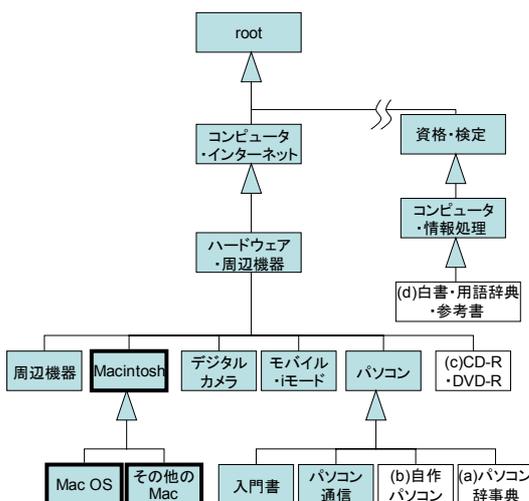


図2 Amazon のカテゴリ階層構造

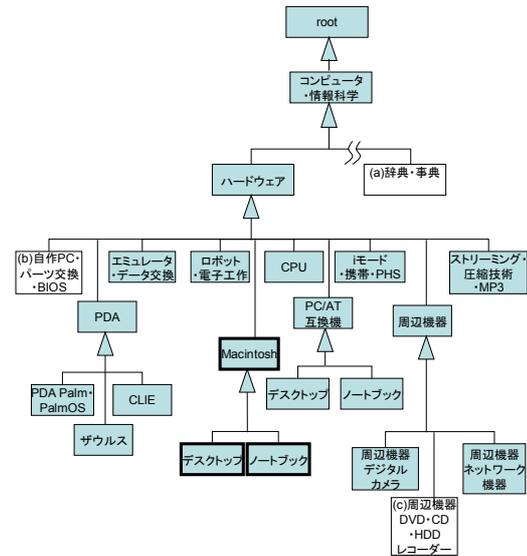


図3 bk1 のカテゴリ階層構造

[1]式から Amazon のハードウェア・周辺機器と bk1 のハードウェアカテゴリについての一致度を求めた結果を表3に示す。

表3 Amazon と bk1 のリソースレベルでの一致度

	カテゴリ数	一致数	一致度
Amazon→bk1	10	4	0.40
bk1→Amazon	16	4	0.25

この結果から、二つの一致度が共にしきい値を超えている場合のみ等価カテゴリとしてマッピングできる。互いの基準カテゴリの下位カテゴリに共通して含まれるものがあると考えればある程度分類の観点は共通しているといえる。

このように一致度はカテゴリに対する分類の観点がある程度等しい場合のみ有効だといえる。互いのサイトの Macintosh カテゴリ について一致度を求めたい場合、基準カテゴリ名から同一範囲を扱っていると推測できるが下位カテゴリの分類の観点が異なるため一致度が求められない。

6.2. リソースレベルでのマッピング

カテゴリレベルでの一致度よりもリソースレベルでの一致度のほうが信頼性は高いといえる。各サイトの最下位カテゴリ同士でマッピングを行った際のリソースレベルでの一致度を確認する。図2と図3で番号が対応する要素間のマッピングの結果を表4で示す。

表4 Amazon と bk1 のリソースレベルでの一致度

		書籍数	一致数	一致度
(a)	Amazon→bk1	38	29	0.76
	bk1→Amazon	38	9	0.23
(b)	Amazon→bk1	41	33	0.80
	bk1→Amazon	41	24	0.58

(c)	Amazon→bk1	30	13	0.43
	bk1→Amazon	30	1	0.03

以上の結果から次の3通りの場合を抽出し、マッピングが行えるかどうかを示す。

(1) 一致度が双方向で高い

商品名から一つのカテゴリに連想しやすい(b)の場合がこれに当てはまる。この場合は等価カテゴリとしてマッピングできる。

(2) 一致度が単方向で高い

(a)の場合がこれに当てはまる。パソコン事典なども一つのカテゴリに連想しやすい。しかし bk1 では事典カテゴリが一つだけ存在するのにに対し、Amazon では類似の事典カテゴリが存在する。このように片方のサイトにおいて商品名から一つのカテゴリに連想しにくい場合が当てはまる。

図2から Amazon では事典に関する2つの類似カテゴリが存在していることが分かる。(a)と(d)の商品名によく含まれる単語を抽出した結果を表5に示す。

表5 各カテゴリでの単語リスト

Amazon		bk1	
カテゴリ	単語リスト	カテゴリ	単語リスト
(a)	辞典, 事典	(a)	事典, 辞典, 用語集, 用語辞典
(d)	用語集, 用語辞典, 白書		

Amazon で(a)と(d)を比較した結果(d)だけに白書という単語が多く含まれる。この分析から(a)と(d)の和集合から商品名に白書が含まれない書籍だけを集めた制約クラスが bk1 の(a)とマッピングできると考えられる。その結果、一致度がどのように変化したのかを表6に示す。

表6 制約クラスを作成しての一致度

	書籍数	一致数	一致度
Amazon→bk1	59	38	0.64
bk1→Amazon	46	17	0.36

この結果、bk1 から Amazon への一致度は上がったが Amazon から bk1 への一致度は下がった。これはプロパティの値の選別においてマッピングできる条件が絞り込めてなかったことが原因と考えられる。

(3) 双方向から見て一致度が低い

(c)の場合がこれに当てはまる。CD-R に関連する書籍が含まれるカテゴリであるがその語句から複数の使用目的が推測できるため人によって分類する基準が異なるといえる。このように二つのサイトで共に商品名からひとつのカテゴリを連想しにくい場合が当てはまる。この場合はカテゴリ名からは二つのカテゴリは等価であると判断できないといえる。

6.3. カテゴリマッピングの評価

カテゴリに含まれる商品情報を見ることで分類の観点の

違いを発見することができた。リソースレベルでの一致度の3通りの結果から、商品の扱う範囲を特定しやすいカテゴリを作るためにはカテゴリ名から判断してそのカテゴリが対象とする「もの」と役割が明確になっている必要があることが分かった。

7. 考察と今後の課題

本研究では UML を用いてカテゴリマッピングを行った。OWL においてはクラス構造の記述の順番は問わないため OWL 文書のままでは全体の構造が把握しにくく構築に手間がかかる。UML では全体の構造の理解が容易であるため UML でオントロジ構築を行うことは有用だといえる。

次に、カテゴリマッピングを行った際の課題を述べる。

(1) データ管理方法

リソースはカテゴリに比べ膨大な数であり日々変化するものである。管理にコストがかかるため機械的にリソースを抽出する仕組みが必要となる。

(2) 制約クラス作成の改善

マッピングの際に適切な制約クラスを作成するために、商品名を見てカテゴリの分類観点の分析方法を改善する必要がある。

8. まとめ

本研究ではオントロジの構築における問題点としてモデルの表現の難しさに焦点を当てた。その解決方法として UML を利用したオントロジの構築方法を提案した。実際の電子商取引サイト間でオントロジを用いたカテゴリマッピングを行うことで UML 利用の利点について確認した。

参考文献

- [1] G. Schreiber, A UML Presentation Syntax for OWL Lite, 2002, <http://hcs.science.uva.nl/usr/Schreiber/docs/owl-uml/owl-uml.html>.
- [2] 神崎正英, セマンティック・ウェブのための RDF/OWL 入門, 森北出版, 2005.
- [3] 長瀬嘉秀他, 独習 UML, 翔泳社, 2005.
- [4] AIDOS, オントロジ技術入門, 東京電機大学出版局, 2005.
- [5] 斎藤信男他, セマンティック Web 入門, オーム社, 2005.
- [6] 伊藤英毅他, 概念体系間の比較と重ね合わせによる関連発見, <http://www-kasm.nii.ac.jp/papers/takeda/99/sigfai00hideta-i.pdf>.
- [7] Amazon, <http://www.amazon.co.jp>.
- [8] bk1, <http://www.bk1.co.jp>.