

高度コミュニティ空間におけるオントロジ言語 OWL を用いた意味層の構築 Formulation of Semantic Layer of Advanced Community Space with OWL

渡邊 優作† 富井 尚志‡
Yusaku Watanabe Takashi Tomii

1 はじめに

近年、ウェブ上での分散的な知識を相互に結びつけるために、情報の詳細な論理表現としてオントロジが用いられている。それに伴い、オントロジを「概念化の明示的な規約」としてマシンリーダブルに表現するため、OWL に代表されるようなウェブ・オントロジ言語が開発され、オントロジは今後、一般的に普及するようになると考えられる[1]。

一方で、我々はオントロジを用いて、仮想空間や実世界などの空間情報を統合し、同じコミュニティの人間が概念を共有する仕組みである、高度コミュニティ空間[2]を提案してきた。この高度コミュニティ空間はデータベースを、概念を蓄積する意味層、インスタンスを表現するマルチメディアデータ層、それらを繋ぐ存在エンティティ層の三層に分けて構築される。これらを独立して管理することで、概念の構造化を可能にし、概念を蓄積する意味層において一般に複雑な構造を持つオントロジの表現を実現している。しかし、前述のセマンティック・ウェブの発展により、オントロジの再利用性が向上したことから、今までの枠組みを越えて、異なるコミュニティでの概念共有が可能になると考えられる。

そこで、本研究ではこのコミュニティを超えた概念の共有を目的とする。この目的に対して本論文では、二つの手法を提案する。一つは、高度コミュニティ空間の意味層に蓄積されているオントロジを OWL で記述可能にし、また OWL で記述されたオントロジを読み込み可能にする。これにより、異なるコミュニティで構築されたオントロジを再利用し、コミュニティを超えた概念共有が可能になると考えられる。二つ目は、二つのオントロジ間にどのクラスが相手のクラスと一致しているのかという対応付けを生成する。これにより、対応付けを通して異なるコミュニティのインスタンスを参照できるようになり、異なるコミュニティとの概念の共有が果たせると考えられる。しかし、この互いのクラスが等しいことを示す対応付けの生成は、以下の二つの問題がある。

- どのクラスが一致しているかをコミュニティごとに指定しなければならない。
- オントロジの複雑性により、手動で対応付けを生成することが困難。

まず一つ目の問題に対して本論文では、セマンティック・ウェブ上に既に定義されたオントロジを媒介にすることによって解決する。二つ目の問題に対しては、対応付けを計算機によって自動的に生成することで解決に導く。この対応付けの自動生成は、オントロジが構造を持つことに着目し、クラス間の関係から二つのオントロジのクラス間

に、Similarity Flooding 法[3][4]を用いることで実現される。本論文では、これらの二つの手法を用いて、高度コミュニティ空間の社会アプリケーションである CONSENT[5]に Ontology Editor を実装、評価することでその実現性を示す。

2 背景

2.1 ウェブ・オントロジ

World Wide Web は人間が視覚的に閲覧するのに魅力的な情報を発信する方向で発展してきた。近年ではその発展として情報にマシンリーダブルなセマンティクスを与えることにより、意味情報によって統合できる WWW 環境である、セマンティック・ウェブの開発が進められている。この手段として、情報の詳細な論理表現としてオントロジが用いられている。オントロジとは一般的に人間の知識を扱う分野、手法を指すが、特に計算機を用いた知識表現としてのオントロジは「概念化の明示的な規約」すなわち「ある分野の知識を計算機で処理できるように明示的、論理的に記述し、その知識の共有と再利用を可能にする」ものと言える[1]。

このオントロジを記述する言語として World Wide Web Consortium によって Web Ontology Language (以下 OWL) が勧告[6]となった。OWL は RDF と呼ばれるデータモデルに基づいて開発され、要素とそれを結ぶ矢印によって表現できる。このように、要素と矢印の表現の最小単位を、その主語、述語、目的語の三項関係からトリプルと呼ぶ。トリプルには個々に意味を付加することができ、OWL や RDF ではこのような意味を構築する手段としての基本的な語彙を用意している。表 1 に、その語彙の一部を挙げる。このような知識を記述する言語が普及することで、様々なコミュニティでオントロジを利用したアプリケーションが普及すると考えられる

表 1 OWL/RDF の語彙

語彙	意味
rdfs:Resource	RDF が表現する全て
rdfs:subClassOf	主語は目的語のサブクラス
rdfs:subPropertyOf	主語は目的語のサブプロパティ
rdfs:domain	プロパティの定義域
rdfs:range	プロパティの値域
owl:Class	OWL のクラスを表すクラス
owl:ObjectProperty	個体値型プロパティを表すクラス
owl:sameAs	同一の個体を示す
owl:inverseOf	反対の関係を示す

2.2 高度コミュニティ空間

我々はコミュニティごとに概念を明示化し、空間データに対応付けて管理・共有する高度コミュニティ空間を提案

† 横浜国立大学大学院環境情報学府
情報メディア環境学専攻
‡ 横浜国立大学大学院環境情報研究院

している。この高度コミュニティ空間は実装基盤としてデータベースを用い、その概念データモデルは三層に分けて構築される。この三層は、コミュニティがもつ概念と概念間の関係を蓄積する意味層、共有対象の空間から得られるあらゆるデータを蓄積するマルチメディアデータ層、その二つの層を結びつけ、マルチメディアデータが示す概念が空間に存在していることを示す存在エンティティ層 (EE層) で定義する[2]。図1に三層構造モデルの要素とそれらの関係を例示した。

本研究の対象である意味層は、対象世界をモデル化したとき、実体を持たない抽象化された概念と概念間の関係が蓄積される層となる。ここで注意すべきことは、意味層を構成する要素は抽象化された概念を表すという部分である。例えば、「机」「本」「置く」といった概念を意味層の要素として蓄積し、さらに「本を机に置く」といった概念も同時に蓄積したとする。このとき、このデータベースが表現可能な空間は、「本を机に置く」ことができる空間であり、実際に「本が机に置かれている」空間を表現するわけではない。この「本」と「机」が存在し、「本が机に置かれている」という事実は EE 層の要素に蓄積することで表現される。

このように抽象化された概念を対象世界に実在する物体や状態を切り分けて構築することにより、利用者コミュニティにとっての「常識」をオントロジとして独立に構造化し、記述できる。

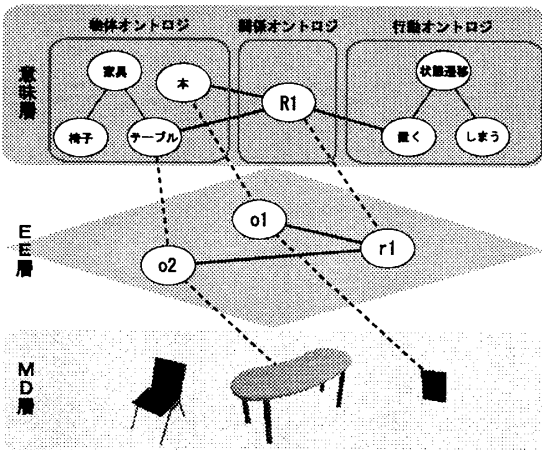


図1 三層構造モデル

3 異なるコミュニティ間の概念共有

先行研究[5]では EDR 電子化概念辞書のような一般的なオントロジを用いることで、概念の共有を果たしていた。しかし、有用なオントロジはコミュニティに大きく依存するため、コミュニティ毎のオントロジ構築は必須である。このオントロジ構築は高度コミュニティ空間の構築における最大のコストとなる。また、コミュニティごとの概念の共有を目的としていたため、異なるコミュニティ間の概念共有を考慮していない。そこで、本研究では、以下の二つの手法を提案することにより、これらの問題を解決する。

3.1 概念の OWL による記述

一つ目は意味層に蓄積された概念とその関係を OWL で記述する。これにより、異なるコミュニティとのオントロ

ジの共有や統合が可能になると考える。この OWL での記述を実現するためには意味層のスキーマと OWL の語彙との対応をとらねばならない。つまり、意味層に構築した概念とその概念間の関係を OWL の語彙によってどのように表現するかを一意に決めなければならない。一意に決めることにより、意味層と OWL 間で情報を落とすことなく変換可能となる。

この意味層と OWL の語彙の対応を示すために、図2に1つの意味層の概念スキーマの例を示す。このスキーマでは意味層に蓄積する概念を、物体に関する概念である物体概念と行動に関する概念である行動概念に分けている。それらの関係を物体概念同士の関係、行動概念同士の関係、物体概念二つと行動概念一つの三種類で対象世界をモデル化している。例えば、「物体概念」エンティティは「本」、「机」、「引き出し」、「家具」などであり、「机が一部として引き出しを持つ」、「机は家具の子である」といった関係が「物体概念間の関係」エンティティである。また、物に対する行動として「本を机に置く」といった概念が「可能行動概念」エンティティとなる。表2にこの例における意味層のエンティティと OWL の語彙との対応の一例を示す。

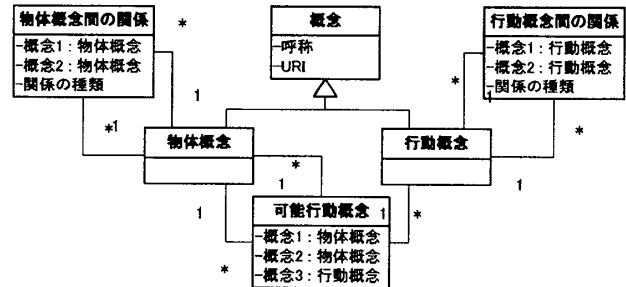


図2 意味層スキーマ例

表2 エンティティと OWL の語彙

エンティティ	OWLの語彙
物体概念	owl:Class
物体概念間の関係	rdfs:subClassOf
行動概念	owl:ObjectProperty
行動概念間の関係	rdfs:subPropertyOf
可能行動概念	owl:ObjectProperty
可能行動概念と物体概念の関係	owl:domain owl:range
可能行動概念と行動概念の関係	owl:subPropertyOf

3.2 対応付けの生成

二つ目の手法として、二つのオントロジ間に対応付けを生成することにより、他のコミュニティと概念の共有を図る。この対応付けとは、二つのオントロジのどのクラスが一致しているかを示すものとする。

しかし、標準化された概念体系がない対象世界において、個別に記述された概念間の対応付けをとることは容易ではない。それには以下の二つの理由が挙げられる。一つは概

念が等しいかという判断は基本的に人間しかできないという理由、もう一つはオントロジが持つ複雑性という理由である。以下では、これら二つの問題とそれに対する解決手法を詳しく述べる。

ウェブ・オントロジの利用

概念と概念が等しいと判断するためには、それらの概念に対する詳細な表現ではなく、二つの概念が同じものを示すという、概念間に対する明示的宣言が必要となる。しかし、異なるコミュニティではお互いのオントロジに対して通常不可知であるため、このような明示的宣言をあらかじめ付加することができないという問題がある。

そこで本研究では、セマンティック・ウェブで記述されるオントロジを媒介として利用することにより、対応付けを生成する。例えば、クラス A とクラス B が共にセマンティック・ウェブで既に定義されているクラス C に対して等しいと記述されるならば、クラス A とクラス B が等しいという対応付けが付加できる。これは、意味層の要素である概念に、それが等しいとするウェブ・オントロジの URI を属性として付加することで実現できる。URI はウェブ・オントロジのクラスやインスタンスに対して一意であり、この属性が等しいことは概念が等しいことを示す。このような媒介に利用するオントロジとして Word Net[7]などが挙げられる。

対応付けの半自動生成

対応付けの生成を妨げる二つ目の問題として、オントロジの複雑性が挙げられる。一般的に概念体系を表現するオントロジはその要素である概念の数が膨大であり、構造が複雑である。よって、この対応付けを人の手で行うことは現実的でない。

そこで本研究では、オントロジ間の対応付けの生成を半自動的に生成することにより、複雑な構造を持つオントロジ間の対応付けに対する人間の負担を軽減する。概念体系を表現するオントロジはその構造を図3のようにグラフで表現することができる。グラフは楕円で示される要素と、概念間の関係を表す矢印の集合で構築される。左右のグラフがそれぞれ別のコミュニティのオントロジを表現していて、異なるオントロジにおいて対応付けをとるとは、破線で示すように、左右のオントロジにおいてどのクラスが対応しているかを明示化することである。

このオントロジがグラフで表現されるという特徴を用いて、対応付けの自動生成には Similarity Flooding 法[3][4]を用いた。この手法は同じ関連を持つクラスは互いに類似している可能性が高いという根拠に基づく。例えば、あるオントロジのクラス A が異なるオントロジのクラス B と同じであるという対応付けが初期に与えられたとする。どちらのクラスにも is-a 関係で結ばれた子を持つとするならば、その子のクラスは A と B が同じという関連から類似している可能性が高いと考えられる。このように二つのグラフの構造から、どの要素が対応しているかを判断する手法である。

この対応付けの自動生成の手法は、初期に与えるマッチングの正確さと数が出力される結果に影響を及ぼす。当然、正確でより多くの初期マッチングを与える方がよい出力を得られる。本研究ではこの初期のマッチングとして前節の

URI 属性を使用することにより、より確からしい結果を得られるように工夫した。

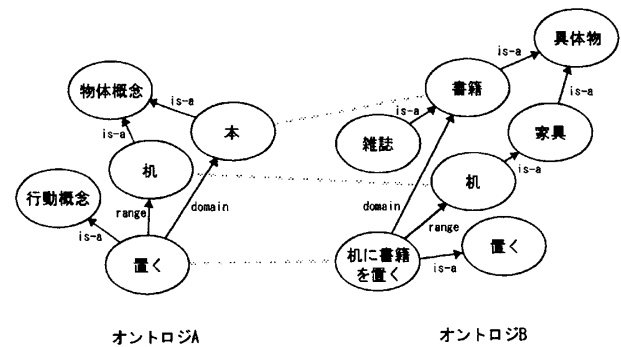


図3 対応付けの生成

4 Ontology Editor の実装

この章では3章で説明した提案手法に基づいて、高度コミュニティ空間[2]の意味層を構築するツールとして Ontology Editor を社会アプリケーション CONSENT[5]に実装した。Ontology Editor は以下の機能を有する。

- 高度コミュニティ空間の意味層の要素を編集する機能。
 - 構築された意味層の要素を OWL で記述する機能。
 - 二つのオントロジ間に対応付けの集合を生成する機能
- 実装した Ontology Editor のアーキテクチャ図を図4に示す。

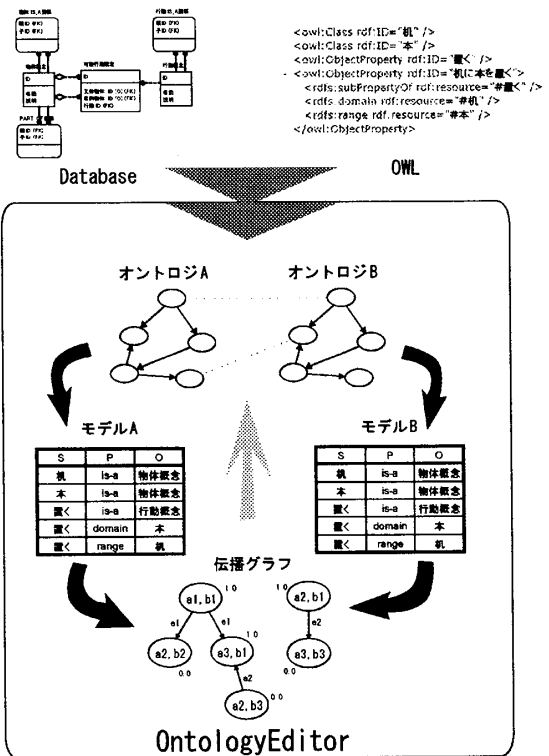


図4 アーキテクチャ図

5 Ontology Editor の評価

この章では前章で述べた、Ontology Editor における対応付けの半自動生成を評価する。まず、自動生成の実現性を評価するため、小規模な二つのオントロジに対して対応付

けの生成を行った。正解セットと初期に与えた対応付けを表 3(a)に、自動生成された対応付け結果の一例を表 3(b)に示す。(a)の灰色の部分今回与えた初期対応付けとなる。初期対応付けは、3.2 節で付加した URI 属性が等しいオントロジの組である。例えば、表 3(a)でオントロジ 1 の「本」、オントロジ 2 の「書籍」というクラスには「http://xmlns.com/wordnet/1.6/Book」という等しい URI が付加されている。この結果により、初期の対応付けをある程度与えるだけで、期待する正解セットに近い形で対応付けが生成できることを示せた。

次に、生成の時間を評価する。オントロジは要素が多く、複雑な構造を持つことは前述したが、本研究の提案手法の実用性評価として、実装した対応付けの自動生成が要素数によってどのくらいの時間を必要とするかを考察する。実装したツールを用いて、この対応付けの自動生成に要した時間をオントロジの要素数に応じてグラフにしたものを図 5 に示す。縦軸はトータルに対応付け処理時間を示し、横軸は要素の数を示す。本研究での手法は Similarity Flooding に基づいており、対応付けの自動生成実行におけるオーダはオントロジの要素に対して多項式時間となる。図 5 に見られるように、ある程度の規模のオントロジになると、自動生成の実行時間は無視できない。しかし、対応付けの生成は頻繁に行われるわけではなく、異なるコミュニティとの概念共有が必要となったときに必要であり、その対応付けを毎回生成する必要はない。よってこの結果から、小・中規模のオントロジに対して、対応付けの自動生成の実用性が示せたと言える。

表 3 対応付け結果

(a)正解セットと初期対応付け		(b)自動生成	
オントロジ1	オントロジ2	オントロジ1	オントロジ2
root	概念	root	概念
具体物	静物	具体物	静物
家具	家具	具体物	家具
台	家具	台	家具
本	書籍	本	書籍
引き出し	引き出し	引き出し	引き出し
テーブル	テーブル	テーブル	テーブル
机	机	机	机
論文集	論文	論文集	論文
漫画	漫画	漫画	漫画
書類引き出し	大きい引き出し	小物引き出し	小さい引き出し
小物引き出し	小さい引き出し	root	物に対する行動
対物行為	物に対する行動	対物行為	移動
対人行為	人に対する行動	対物行為	開け閉め
借りる	借りる	置く	置く
貸す	貸す	しまう	片付ける
置く	置く	取り出す	取り出す
しまう	片付ける	閉める	閉める
取り出す	取り出す	開ける	開ける
閉める	閉める		
開ける	開ける		

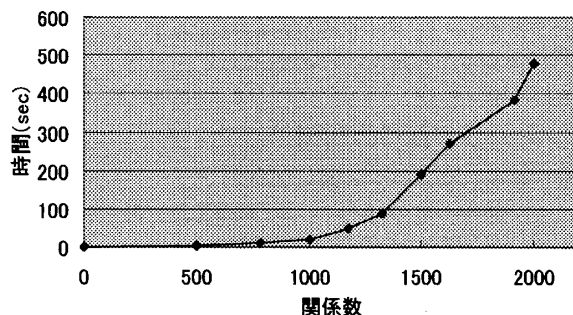


図 5 対応付けの時間

6 まとめ

本論文では、高度コミュニティ空間において異なるコミュニティとの概念共有を目的とし、それを実現する手法を提案した。また、提案手法に基づいて、Ontology Editor の実装を行った。この論文で挙げた提案手法を以下にまとめる。

- 意味層に蓄積された要素の OWL での記述。
- 異なるコミュニティでのオントロジ間の対応。

これらによって、異なるコミュニティのインスタンスが参照可能となり、コミュニティ間の概念共有が可能となったと考える。

今後の課題として、データベースと OWL の語彙の対応の強化が挙げられる。これは OWL が根幹とする RDF が三項関係のシンプルなデータモデルであるため、n 項関係を表現する際、不要に複雑な記述が必要となるからである。これを防ぐには RDF のモデルに対応したデータベースのスキーマ構築が不可欠であると考えられる。

謝辞

本研究の一部は文部科学省科学研究費補助金(課題番号 17700097)の支援による。

参考文献

- [1] 神崎正英, “セマンティック・ウェブのための RDF/OWL 入門”, 森北出版株式会社, Jen, 2005
- [2] 富井尚志, “マルチメディアデータベースに基づく高度コミュニティ空間の実現”, 電子情報通信学会誌, Vol.89, No.6, pp.511-517, June, 2006
- [3] S. Melnik, H. Garcia-Molina, E. Rahm, “Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching”, proceedings of the 18th International Conference of Data Engineering (ICDE 2002), pp. 117-128, 2002
- [4] S. Melnik, E. Rahm, P. A. Bernstein, “Rondo: A Programming Platform for Generic Model Management”, proceedings of the ACM SIGMOD 2003, pp. 193-204, 2003
- [5] 小川悌知, 賀来健一, 渡邊文隆, 佐渡山英史, 富井尚志, “概念共有環境 CONSENT における行動パターンの取得”, データ工学ワークショップ論文集(DEWS2006), 4B-i9, Mar, 2006
- [6] OWL 勧告, <http://www.w3.org/TR/>
- [7] Word Net, <http://wordnet.princeton.edu/>