

Linked Open Data を用いた 生物規範工学オントロジーの大規模化の試行

A Trial to Expand a Biomimetics Ontology using Linked Open Data

多田恭平¹ 古崎晃司¹ 來村徳信² 溝口理一郎³ 駒谷和範¹

Kyohei Tada¹ Kouji Kozaki¹ Yoshinobu Kitamura² Riichiro Mizoguchi³ Kazunori Komatani¹

¹ 大阪大学産業科学研究所

¹The Institute of Scientific and Industrial Research (ISIR), Osaka University

² 立命館大学

² Ritsumeikan University

³ 北陸先端科学技術大学院大学

³Japan Advanced Institute of Science and Technology

Abstract: It is required to develop biomimetics database that supports engineers to develop new products inspired by biological functions. Biomimetics ontologies are the foundation of the biomimetics database. There are two methods to expand biomimetics ontologies. We expanded them using technical document analysis so far. In this paper, we especially explain a method using Linked Open Data. We particularly discuss a trial to expand the ontology using DBpedia Japanese.

1. はじめに

生物規範工学は、生物がもつ優れた機能を模倣し、技術開発や物作りに活かすことを目指している。ここでは、昆虫や魚類、鳥類といった生物が有する多様性に関する知識を、製品開発者が様々な観点から検索できるバイオミメティクス・データベースの構築が求められている[1]。このデータベースは、新たな技術を開発しようとする工学研究者に対し、技術革新の着想につながる発想支援の実現を目指して開発が進められている。

具体的には、工学研究者が求める「機能」から、その機能を実現している可能性のある「生物」を示唆するなどの利用が想定されている。その際に、より発想を刺激するためには、機能と生物の直接的な関係のみならず、機能の達成関係、生物の生態環境と関連する機能など、様々な関係を用いた検索が重要となる。

そのような発想支援につながる検索を実現するために、生物規範工学に関する様々な知識を体系化した、生物規範工学オントロジーの構築が進められている[2]。そのオントロジーを対象として、利用者の関心に応じた観点から定義された関係をたどる技術（オントロジー探索ツール）[3]を適用することで、

意味構造に基づいて、多種多様な生物の知識をさまざまな観点から検索可能となる。

ここで、発想支援につながる検索結果をより多く得るためには、検索の対象とする生物規範工学オントロジーの大規模化が必要とされる。すなわち、そのオントロジーが、より多くの概念を含み、それらが意味のあるつながりで結ばれることにより、より多様な検索結果が得られる。

そのようなオントロジーを構築するために、本研究では多種多様な生物に関する情報源から、必要な概念や関係を半自動的に獲得することで、オントロジーを大規模化する方法の開発を進めてきた[4]。本研究における半自動的にオントロジーの大規模化は、専門文書の解析と Linked Open Data (LOD) の利用という2つのアプローチから行っている。本稿では、そのなかでも Linked Open Data を利用した大規模化について述べる。

以下、2章では本研究で提案するオントロジーを大規模化する手法について述べる。3章では、Linked Open Data を利用したオントロジーの大規模化を、DBpedia Japanese を用いて試行した結果について述べる。4章では今後の課題を述べると共に、本論文を



図 1 目指しているオントロジーの記述内容の例[5]

総括する。

2. オントロジー大規模化

2.1 目指すオントロジー

本研究で大規模化の結果として得ることを目指している、生物規範工学オントロジーの一部を図 1 に示す[5]。オントロジーは概念と概念間の関係を定義することで構築されるが、概念間の関係については一般に、大きく 2 種類に分けて考えることができる。is-a リンクで表される is-a 関係と、スロットで表されるその他の種類の関係である。

is-a 関係は、その概念（クラス）間の一般-特殊関係、すなわち上位クラス（上位概念）と下位クラス（下位概念）の関係を表す。図 1 の例では、「カタツムリ」の上位クラスが「動物」であることが is-a 関係で表されている。

一方、「カタツムリ」の“生態環境”は「湿地」であることや、「カタツムリ」の“特徴的機能”に「防汚」があるといったその他（is-a 以外）の関係はスロットを用いて定義される。スロットの上部に書かれた“生態環境”などのラベルが関係の種類を表している。これらの関係を用いることで、例えば「防汚」機能をもつ生物を検索した際に、「カタツムリ」を検索結果として出力するといった検索システムが構築できる。

この生物規範工学オントロジーを用いて、「防汚、抗菌塗料」を実現するために参考にできる機能をもつ「生物」を検索した結果を図 2 に示す。この検索結果の図から、概念同士が「関連機能」や「特徴的構造」といった概念間の意味のラベルがつけられた関係で結ばれていることと、それぞれの概念が上位クラスの種類が色の違いによって示されていることが分かる。意味のある関係は、オントロジー上で「防

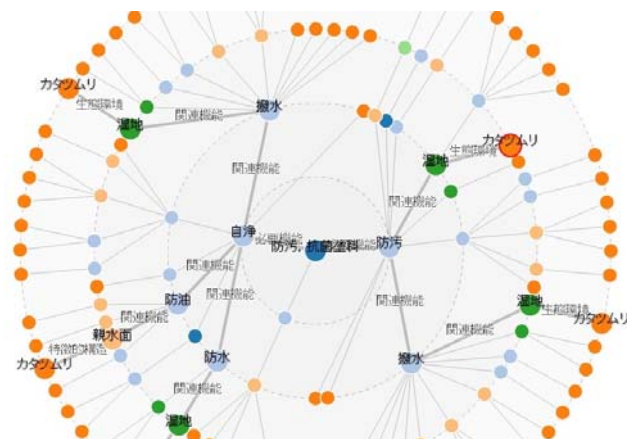


図 2 生物規範工学オントロジーを用いた検索結果の例 (<http://biomimetics.hozo.jp/>にて検索)

汚、抗菌塗料」と関係のある概念を、「生物」の下位クラスに到達するまで再帰的に探索することで得られた、概念間の関係の系列（パス）を示している。ここで、“関係がある概念”の探索には、スロットで定義された関係が用いられる。一方、is-a 関係は各パス上に現れる概念の上位クラスの判定に用いられ、図 2 では、生物、機能、生態環境といった上位クラスの種類ごとに色分けして表示される。

すなわち、オントロジー探索ツールを用いて生物規範工学オントロジーの検索を行うには、is-a 関係とスロットで定義されたその他の関係の双方が必要である。よって、本研究では、それらの関係に注目してオントロジーの大規模化を行う。ただし is-a 関係については、「生物」、「機能」、「構造」といった主要な上位クラスの判定のみに用いられる。そのため、本研究で行う生物規範工学オントロジーの大規模化は、追加される概念の上位クラスはそれらの何れに相当するかが分かればよいものとする¹。

2.2 オントロジーの半自動大規模化の手法

と手順

生物規範工学における、より一般的な知識をオントロジーに追加するために、本研究では Linked Open Data を利用した大規模化を行っている。その手順は以下の 3 ステップからなる。

(1)オントロジーに追加する概念の候補の選択

情報源となる LOD から、生物規範工学オントロジーに追加する概念の候補となる単語を選択する。

(2)オントロジーに追加する概念の上位クラスの同

においては、1 段の is-a 階層のみでも必要な検索が行える。

¹ 通常のオントロジーでは、is-a 関係による多段のクラス階層（is-a 階層）が形成されるが、本研究の目的

定

追加候補の単語をオントロジーへ追加する箇所、すなわち、候補概念の上位クラスを同定する。

(3)その他の関係の種類と同定

同定した上位クラスの情報を用いて、追加概念と他の既存概念との間の、is-a 関係以外の関係の種類を同定する。

また、生物規範工学特有の知識を獲得する方法として、自然言語処理技術を用いて専門書を解析することを提案しているが、こちらについても同様に上述した3ステップから大規模化が行える。

2.3 基となるオントロジー

生物規範工学オントロジーを半自動的に大規模化するにあたり、すべての概念を自動処理で取得することは困難である。そこで、大規模化作業の基となるオントロジーを手動であらかじめ構築することとした。これまでに筆者らは、昆虫を対象としたバイオメティクスの専門書である「昆虫ミメティクス」[6]に付録されているキーワード索引を用いて、基となるオントロジーを構築した。キーワード索引とは、専門書の著者が重要であると判断した語をキーワードとして選出し、索引化したものである。

先行研究において、キーワード索引に挙げられている単語 657 語を、17 の上位クラスに手動で分類することで、オントロジーを構築した。このオントロジーで導入した 17 の上位クラスを最上位クラスと呼ぶ[4]。そして、それらの最上位クラスの中から、手動で構築されたオントロジー[5]を参考にして、生物規範工学オントロジーにおいて重要であると考えられる最上位クラスを 5 つ選定した（機能、行為、構造、部位、物質）。

本研究ではこの 5 つの最上位クラスを上位クラスとする概念を基となるオントロジーに追加することを、大規模化の当面の目標と設定した。以降では、この基となるオントロジーを利用したオントロジーの大規模化について議論する。

3. Linked Open Data を利用したオントロジー大規模化の試行

3.1 利用する LOD

² DBpedia に含まれるデータ（リソースと呼ばれる）には、クラスとインスタンスの双方が含まれるが、本研究で大規模化の対象としている生物規範工学オントロジーを用いた検索においては、それらの明確

前章で述べたように、Linked Open Data を利用した大規模化は、基となるオントロジーを利用し 3 つのステップから行う。利用する LOD について、本論文では DBpedia Japanese を利用した大規模化の試行について議論する。DBpedia Japanese は、Wikipedia の日本語版のデータを用いて構築された LOD であり、Wikipedia の各記事に関するデータが LOD として公開されている。日本語の LOD としては最も規模が大きいものであるため、多くの研究で利用されている。

本研究では、生物規範工学に関わる、より一般性の高い概念の拡充と、提案手法である LOD を用いたオントロジーの大規模化の基本的な性能を評価するために、DBpedia Japanese を用いた大規模化を試行した。その中でも本稿では、大規模化の 3 つのステップのうち、(1)オントロジーに追加する概念の候補の選択と(2)オントロジーに追加する概念の上位クラスの同定について述べる。以下、それぞれのステップ毎に行った手順とその結果について述べる。

3.2 追加する概念の候補の選択

オントロジーの大規模化を行うにあたり、発想支援に有用な概念間の関係を優先的に追加するため、基となるオントロジーで定義された概念と関係がある概念を優先的に追加するという方針をとった。基となるオントロジーの 657 概念のうち、DBpedia Japanese にデータが存在するものは 287 個（43.7%）であり、それらの概念数を、本研究で注目している 5 つの最上位クラス毎にまとめたものを表 1 に示す。

これら 287 個の概念に関して、DBpedia Japanese 上での関係（プロパティ）が定義されている概念²（データ）を、追加する概念の候補とする。今回の試行では、定義されている関係の数が多ことから、Wikipedia 上の各記事内でハイパーリンクされている他の見出し語を表している dbpedia-owl:wikiPageWikiLink という関係（プロパティ）を用いることとした。すなわち、基となるオントロジーで定義されていた 287 個の概念と、dbpedia-owl:wikiPageWikiLink によって関係づけられている単語を抽出することで、オントロジーに追加する概念の候補として選択した。

な区別は必要とされていない。よって、本稿においても、クラスとインスタンスの区別については議論しない。

表 1 最上位クラス毎の概念数

最上位クラス	概念数
機能	54
行為	22
構造	9
部位	40
物質	21
5つ以外のクラス	141
合計	287

表 2 最上位クラス毎の上位クラスの数

最上位クラス	上位クラスの数	重複を取り除いた数
機能	766	267
行為	343	50
構造	425	112
部位	407	54
物質	402	88

表 3 最上位クラスを代替することの確からしさの分布

最上位クラス	0~0.05	~0.10	~0.15	~0.20	~0.25	~0.30	~0.35	~0.40	~0.45	~0.50	合計
機能	187	52	17	6	4	1	0	0	0	0	267
行為	0	50	0	0	0	0	0	0	0	0	50
構造	0	0	9	0	0	0	0	0	0	0	9
部位	38	14	2	0	0	0	0	0	0	0	54
物質	0	33	30	12	7	3	1	0	0	2	88

その結果、DBpedia Japanese から 10,017 単語が抽出され、これを追加する概念の候補とした。

3.3 追加する概念の上位クラスの同定

3.3.1 上位クラスの同定の概要

本研究では、前節で選択した追加候補が、どの最上位クラスの下位概念となるかを同定することを目指す。そのために、まず基となるオントロジーの最上位クラスに相当する DBpedia Japanese 上のクラスをみつける必要がある。そこで、DBpedia Japanese 上で、Wikipedia の記事のカテゴリ分類に相当するデータを、クラス階層に代替するものとして用いることとした。

具体的には、DBpedia Japanese にて定義されている 2 種類のプロパティ情報を利用した。1 つ目は、各データの Wikipedia 上でのカテゴリを表すプロパティである `dcterms:subject` である。これを用いることで、そのデータが属しているカテゴリが同定できる。2 つ目は、そのカテゴリの広義語を表すプロパティである `skos:broader` である。これを用いることで、各カテゴリの上位カテゴリを辿ることができる。

本研究では、これら 2 つのプロパティを再帰的に辿って得られるカテゴリの系列を、それらの上位クラスの系列として利用することとした。これにより、追加候補となる概念の上位クラスを求め、それが、

基となるオントロジーの最上位クラスに相当するものとなれば、その追加候補となる概念の最上位クラスが同定できる。

しかし、本研究で注目した 5 つの最上位クラスに直接対応するカテゴリ (クラス) は、DBpedia Japanese には存在しなかった[4]。そこで、基となるオントロジーで「各最上位クラスの下位クラスとして定義されており、かつ DBpedia Japanese に対応するデータが存在する概念」のうち、上位クラスが定義されているものについて DBpedia Japanese 上での上位クラスを 5 段辿り、上位クラスのリストを得た。そして、このリストに含まれる概念を、「各最上位クラスに相当する概念」として用いることとした。すなわち、候補単語の上位クラスがこのリストに含まれているとき、その候補単語は、該当する最上位クラスの下位概念になると同定する。

しかし、これらのリストに含まれる上位クラスの中には、一つの最上位クラスのリストだけではなく複数の最上位クラスのリストに含まれるものも存在する。それらは候補単語の最上位クラスを一つに決めることには利用できないため、各リストの中から複数のリストに含まれる上位クラスを取り除いたものを作成し、これを「各最上位クラスに相当する上位クラス」のリストとした。その結果として得られた、各リストに含まれる上位クラスの数を表 2 に示す。³

³ 例として「共生」は機能と行為のリストに存在した

ため取り除いた。

表4 最上位クラスの同定精度

		追加候補の上位クラスがリストに含まれた数	同定の精度
機能	リストの全概念	1120	19/50(38%)
	確からしさが 0.1 以上の概念	197	5/50(10%)
行為	リストの全概念	148	1/50(2%)
	確からしさが 0.1 以上の概念	0	-
構造	リストの全概念	33	8/33(24%)
	確からしさが 0.1 以上の概念	33	-
部位	リストの全概念	400	18/50(36%)
	確からしさが 0.1 以上の概念	12	12/12(100%)
物質	リストの全概念	552	26/50(52%)
	確からしさが 0.1 以上の概念	314	33/50(66%)

また、各最上位クラスのリストの中で、その下位クラスに、基となるオントロジー中で同じ最上位クラスの下位に相当する概念をより多く含んでいるものは、その最上位クラスを代替することが「より確からしい」と考えられる⁴。そこで、最上位クラス毎の上位クラスのリストにおいて、各上位クラスが「その最上位クラスの下位概念（表1の数）中、いくつの概念の上位クラスになっているか」の割合を計算した。その値を、その上位クラスで最上位クラスを代替することの確からしさとする。その確からしさの分布は表3に示す。

3.3.2 上位クラスの同定精度の検証

提案手法による追加概念候補に対し、前項で導入した「各最上位クラスに相当する上位クラス」のリストを用いて、追加候補単語の上位クラスが、最上位クラスのどれに相当するかを同定した。追加候補単語の上位クラスには、候補単語と DBpedia Japanese 上で dcterms:subject で関係づけられているカテゴリを用いた。

最上位クラスについては、前章で述べた本研究で注目する5つの最上位クラスのみを考え、各同定結果の精度（適合率）を調べた。適合率は、最上位クラス毎に追加候補を50個ランダムに抽出し、同定された最上位クラスが適切であるかを著者が判定した結果を用いて計算した。

この試行において最上位クラスの同定は、「最上位

クラスを代替することの確からしさ」の有用性を調べるために、各上位クラスのリストの全概念を対象にしたときと、リストの中で確からしさが0.1以上である概念のみを対象としたときの2つのパターンで行った。これらの結果を表4に示す。

結果から、部位、物質の2つの最上位クラスについては確からしさを導入することで上位クラス同定の精度が上がっていることが分かる。他の3つの最上位クラスの同定の精度が低いことについては、次のような原因が考えられる。

- ・機能

基となるオントロジーにおいて機能の下位概念として定義されている「超撥水」や「発光」といった概念は、DBpedia Japanese 内で学問を表すクラスの下位概念として定義されている（「超撥水」は「流体力学」の下位概念、「発光」は「光学」の下位概念としてそれぞれ定義されている）。そのため、学問を表すクラスの下位概念にはその学問に関する機能以外の単語が多く定義されており⁵、それが上位クラスの同定精度を下げている原因となっている。

- ・構造、行為

基となるオントロジーにおいて機能や構造、行為の下位概念として定義されている概念は、意味的にお互いの下位概念と似ているため、上位クラスのリスト同士の重複を取り除く段階で行為を表す上位クラスがリストから除外されている。例えば、行為の下位概念である「変態」の上位クラスである48個の

⁴ 物質の上位クラスリストにある「無機化学」は基となるオントロジーの21個の概念のうち6個の概念の上位クラスとなっている。一方、「二糖」は21個の概念のうち1個の概念の上位クラスにしかなくない。よって、「無機化学」のほうが「二糖」より上位クラスを「より確からしく」同定できると考え

られる。

⁵ 例として、「流体力学」の下位概念は182概念定義されているが、機能の下位概念として考えられるものは26概念しかない。他には学問の名前、関係する方程式の名前などが定義されている。

概念は全て他のリストと被っているため除外されている。構造と行為については、基となるオントロジーにて定義されている概念が少ないため、利用できる上位クラスも少なく、これが精度の低い原因となっていると思われる。

4. 本研究の総括と今後の展望

本研究では、生物の多様性に学ぶものづくりを目指した生物規範工学分野において、工学研究者に対して新たな製品開発の発想支援に利用される、生物規範工学オントロジーの Linked Open Data を利用した大規模化を試行した。

今回の試行では、主に追加する概念候補の上位クラスの同定を中心に行った。その結果、今回の同定精度は、2%~100%と、最上位クラスの種類によって、大きな差が見られた。今後、上位クラスの同定精度を向上させるために、

- ・最上位クラスを代替すること確からしさ以外の尺度の導入
- ・追加候補の上位クラスを多段にみる
- ・複数の上位クラスを持つ追加候補の扱いの検討などの、同定手法の改善を検討する。

また利用する LOD について、現状の DBpedia Japanese 以外のものを対象とすることを計画している。具体的には、日本語 Wikipedia オントロジー^[7]⁶、LODAC Species⁷、Web NDL Authorities⁸といった他の LOD を用いることで、まだ同定精度が十分でない機能、構造、行為について、最上位クラスの同定精度を上げることができるかを検証する。

なお、本研究におけるオントロジーの大規模化は、「昆虫」の多様性を模倣した生物規範工学の知識を対象として行ったが、他の生物種についても同様の手法を順次適用していく。

謝辞

本研究の一部は科学研究費補助金 新学術領域(研究領域提案型) 24120002「バイオミメティクス・データベース構築」の助成による。

参考文献

- [1] 下村 政嗣: 生物の多様性に学ぶ新時代 バイオミメティクス材料の新潮流, 科学技術動向 Vol.110, pp.9-28, 2010
- [2] 古崎 晃司, 來村 徳信, 溝口 理一郎: 生物多様性を規範とした材料技術開発支援に向けたバイオミメティクス・オントロジーの試作, 2013 年度人工知能学

会全国大会, 3I3-3, 2013

- [3] K. Kozaki, T. Hirota, and R. Mizoguchi : Understanding an Ontology through Divergent Exploration, In Proc. of 8th Extended Semantic Web Conference (ESWC2011), pp.305-320, Heraklion, Greece, May 29 - June 2, 2011
- [4] 多田 恭平, 古崎 晃司, 來村 徳信, 溝口 理一郎, 駒谷 和範: 概念間の関係に注目した専門文書解析と LOD 技術によるバイオミメティクス・オントロジーの大規模化の試み, 2015 年度人工知能学会全国大会, 1G4-3, 2015
- [5] 鳥村 匠, 來村 徳信, 古崎 晃司, 溝口 理一郎, 駒谷 和範: 生物の機能実現方法に基づく発想支援のためのオントロジー構築とそのガイドラインの提案, 2015 年度人工知能学会全国大会, 2M1-5, 2015
- [6] 下澤 楯夫: 昆虫ミメティクス~昆虫の設計に学ぶ~, NTS, 2008
- [7] 玉川 奨, 森田 武史, 山口 高平: 日本語 Wikipedia からプロパティを備えたオントロジーの構築, 人工知能学会論文誌, Vol.26, N0.4, pp.504-517, 2011

⁶ <http://www.wikipediaontology.org/>

⁷ <http://lod.ac/>

⁸ <http://id.ndl.go.jp/auth/ndla/>