2F1-4

オントロジーと Linked Data に基づくバイオミメティック・データベースの構築

A Development of Biomimetic Database based on Ontology and Linked Data

古崎 晃司[*1] Kouji Kozaki　多田 恭平[*1] Kyohei Tada　來村 徳信[*1] Yoshinobu Kitamura　溝口 理一郎[*2] Riichiro Mizoguchi

[*1] 大阪大学産業科学研究所 The Institute of Scientific and Industrial Research (ISIR), Osaka University

[*2] 北陸先端科学技術大学院大学 Japan Advanced Institute of Science and Technology

For biomimetics research, it is important to develop biomimetics database which enable us to find a huge variety of knowledge across different domains. To realize such database, an interoperability of knowledge between them is necessary. Ontologies clarify concepts that appear in the target domains and contribute to improvement of interoperability. Furthermore, in order to integrate the database with existing databases for biological diversity, linked data technologies are very effective. This article discusses a development of biomimetics database for materials engineering based on ontology and linked data.

1. はじめに

「自然に学ぶものつくり」を目指すバイオミメティクス（biomimetics）研究においては，新たな技術を開発しようとする工学研究者が，生物多様性と適応に関する情報を通した技術革新の着想を得ることのできる，バイオミメティクス・データベースの整備が重要とされる[下村 10]. 本研究では，このような発想支援型のデータベースを開発するにあたり，バイオミメティクスに関する知識を体系化したバイオミメティクス・オントロジーを構築し，「利用者の視点に応じたオントロジー探索技術」[Kozaki 11]および「オントロジー知的探索に用いる概念検索の対象範囲を適切に管理する技術」[古崎 13a]を用いた検索システムの開発を進めている[古崎 13b].

例えば，“材料研究者が求める「機能」から，その機能を実現している「生物（の部位）」を検索する”ことを考えると，工学的な材料と生物の機能を直接的に結びつけるだけでは無く，“「汚泥」を「生活環境」とする生物は「防汚機能」を持つ”といった生物学の知識が必要とされる別の観点からの検索も可能とするようなオントロジーが必要となる. 先行研究で試作したバイオミメティック・オントロジーを用いると，注目する「機能」から「生物種」，「生態環境」，「生物の行動」，「構造」といった，様々な観点からの概念のつながりが見える. バイオミメティック・データベースでは，これらの概念のつながりを利用して，それぞれの概念と対応づくメタデータが付与されたデータを検索することで，オントロジーに基づく検索機構を実現する.

さらに，バイオミメティクス研究においては，生物学と工学にとどまらず，様々な領域の知識をつなぐことで新たなイノベーションの促進が期待されるため，バイオミメティクス・データベースには，既存のデータベースとの相互連携を促進する仕組みが重要となる. この要件を満たすデータベース開発の為に，本研究では， Semantic Web 技術を用いて Web 上のデータを相互に連携(Linking)させることにより新しい価値を生み出そうとする Linked Data 技術[ヒース 13]を利用する.

以下，2 章ではバイオミメティクス・オントロジーの構築とデータベースにおける利用について概説し，3 章では Linked Data を用いたバイオミメティクス・データベース構築についに述べる. 4 章では本研究の現状をまとめると共に，今後の展望について述べる.

連絡先：古崎晃司，大阪大学産業科学研究所 知識科学研究分野，〒567-0047 大阪府茨木市美穂ヶ丘 8-1，TEL:06-6879-8416，kozaki@ei.sanken.osaka-u.ac.jp

2. バイオミメティクス・オントロジーの構築と利用

先行研究において，博物館に所属する昆虫および魚類の研究者から提供された生物種に関する概要説明文書の情報を元に小規模なオントロジーを試作した. 対象とした生物種は，別グループでバイオミメティックの画像データベース用に電子顕微鏡写真が撮影した昆虫 13 種，魚類 12 種である. 本オントロジーは，それぞれの生物の種・科・目の情報に沿った is-a 階層と，生物種毎の「特徴的な機能」，「構造」，「行動」，「生態環境」等の属性定義からなる. オントロジー構築には「法造」[1]を用いており，定義された概念数は 226，属性を表すスロット数は 133 である. このオントロジーを対象としたオントロジー探索により，

・機能 → 生物種 → 構造
・機能 → 生態環境 → 生物種 → 構造
・機能 → 生物の行動 → 生物種 → 構造
・機能 → 構造 → 生物種

といった，様々な観点からの概念間のつながりが探索できることが確認できている[古崎 13b].

しかし，多種多様な生物の特徴を模倣した技術革新につながる発想支援の実現には，より多くの生物種を対象としたオントロジー構築が必要となる. 地球上に存在する数百万種を越える生物種を対象とすることを考えると，オントロジー構築の自動化は必須の課題である. そこで，本研究と並行して，専門文書および Linked Open Data を用いたバイオミメティック・オントロジー大規模化手法の開発を進めている[多田 14].

その結果，バイオミメティクス・オントロジーの大規模化が進むと，上述のようなオントロジー探索によって得られる概念間のつながりの組み合わせが膨大になり，適切な絞り込みが必要となることが想定される. 本研究では，オントロジーの is-a 階層に沿った属性継承の性質を利用することで，オントロジー探索に用いられる概念検索の対象範囲を適切に管理する「多段階展開型検索手法」[古崎 13a]を適用することで，探索範囲の適切な絞り込みを行う.

3. Linked Data を利用したバイオミメティクス・データベースの構築

3.1 生物種情報に関する Linked Data

近年，生物多様性情報に関するデータベースは多数開発さ

[1] http://www.hozo.jp

れており，それらの統合利用が進められている[大澤 14]．Linked Data は Web 上に公開されたデータベースを統合利用する技術として注目されており，ライフサイエンスやオープンガバメントの分野をはじめ，多くの領域で Linked Data 技術に準拠したデータベースが公開されている[1]．

本研究で開発するバイオミメティクス・データベースは，工学の研究者が新たな技術開発の着想につながる生物種の情報を得ることが第一の目的であるため，遺伝子配列など各生物の詳細な情報よりも，生物の生態など概要情報を得ることが重要となる．そのような情報を含む Linked Data として，生物の種名情報を対象とした Lodac Species[2] [南 11]，Wikipeda の情報を抽出することで構築された多くの Linked Data とリンクするハブとして広く利用されている DBpedia の日本語版[3]および英語版[4]，日本語 Wikipedia を元によりリッチな情報を含むオントロジーとして構築・公開されている日本語 Wikipedia オントロジー[5][玉川 11]を対象として，バイオミメティクス・データベースへの利用を検討する．

3.2 Linked Data 利用に向けた予備的検証

まず，各 Linked Data が必要な情報を含んでいるかを予備的に検証するために，昆虫および魚類の研究者から提供された，バイオミメティクスの画像データベースに写真を格納する生物の，目・科・種を対象としたデータ計 46 種が各 Linked Data に含まれているかを調べた結果を表1に示す．各データの有無の判定は，各生物の目・科・種の和名による文字列の完全一致検索で該当データが取得できるか否かで行った．

この結果より，Lodac Speices は検証対象とした全データが各生物種の「和名」として含まれていることが分かり，各和名に対応する「学名」など種名に関する情報を取得する際に有効であることが分かった．DBpedia については，日本語版では対象データの 9 割以上がカバーされているが，英語版ではカバー率がその半数にとどまっている．これは，日本語と英語のデータの対応が完全ではないため，和名では英語版 DBpedia の該当データが正しく取得できないためと思われる．ただし，Lodac Speices から取得した「学名」等を利用することで，英語版から該当するデータを取得することができる可能性はあるので，今後，検証したい．

日本語 Wikipedia オントロジーのカバー率が，DBpedia の日本語版よりも小さくなるのは，日本語 Wikipedia オントロジー構築の過程で利用されているデータの補完・修正処理等が何らかの影響を与えている可能性が考えられる．日本語 Wikipedia オントロジーと DBpedia の日本語版は，共に，日本語 Wikipedia を元に構築されているので，本来は該当データの存在数は一致するはずである．この差については，日本語 Wikipedia オントロジーと DBpedia 日本語版の間のマッピング情報を参照するなどして，原因をより詳細に検討したい．

続いて，これらの Linked Data から得られる情報が，バイオミメティック・データベースに有用であるかの予備的検証を行うために，種名情報のみを対象としている Lodac Speices 以外で該当データの存在数が最も多い DBpedia 日本語版を対象として，各生物種の情報を取得した．対象とした生物種は上述の昆虫，魚類に加えて，鳥類を加えたものである．

[1] http://lod-cloud.net/
[2] http://lod.ac/species/
[3] http://ja.dbpedia.org/
[4] http://dbpedia.org/
[5] http://www.wikipediaontology.org/

表1 バイオミメティクスの画像データベースに格納される生物の目・科・種のデータが Linked Data に含まれる数

対象とする Linked Data	該当データの存在数	該当データの存在割合(%)
Lodac Species	46	100
DBpedia 日本語版	43	93.5
DBpedia 英語版	19	41.3
Wikipedia オントロジー	30	65

表 2 DBpedia 日本語版から取得した情報の例（エドアブラザメの例）

Wikipedia における記事概要	エドアブラザメ Heptranchias perlo（江戸油鮫、英: Sharpnose sevengill shark）は、カグラザメ目カグラザメ科に属するサメ。本種のみでエドアブラザメ属 Heptranchias を形成する。@ja
界	動物界
門;亜門	脊索動物門;脊椎動物亜門
綱;亜綱	板鰓亜綱;軟骨魚綱
目	カグラザメ目
科	W:Heptranchias;カグラザメ科
属	W:Heptranchias
Wikipedia における関連記事の項目名	脊索動物門;W:Vertebrata;W:Chondrichthyes;W:Heptranchias;W:Chordata;Category:カグラザメ目;サメ;鉤;板鰓亜綱;W:Animalia;ファイル:Teeth_of_sharpnose_sevengill_shark_(Heptranchias_perlo.jpg;櫛;W:Hexanchidae;脊椎動物亜門;W:Hexanchiformes;軟骨魚綱;1788年;エイ;頭足類;動物界;カグラザメ目;カグラザメ科;甲殻類;ファイル:Sharpnose_sevengill_shark_(_Heptranchias_perlo_).jpg;硬骨魚類;ファイル:Heptranchias_perlo_distmap.png;W:Elasmobranchii;胎生

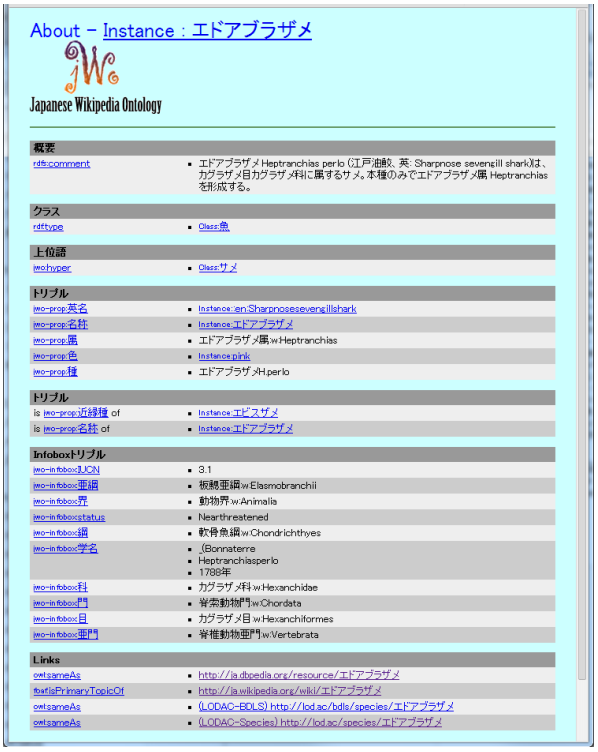

図 1 日本語 wikipedia オントロジーから取得できる情報の例（エドアブラザメ）．

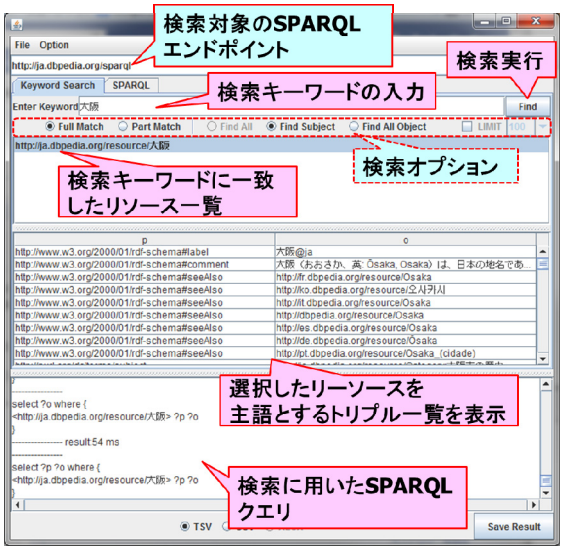

図 2 簡易 SPARQL ツールの実行画面例.

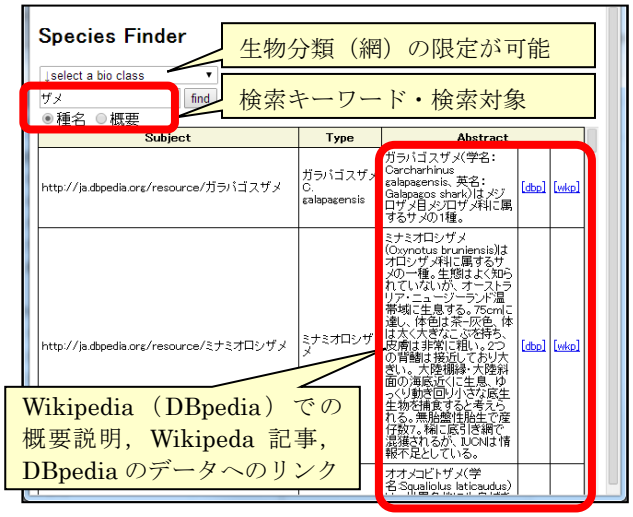

図 3 簡易 SPARQL ツールの実行画面例.

DBpedia 日本語版から取得できる主な情報は，「界」，「門」，「亜門」，「綱」，「亜綱，「目」，「科」，「属」といった生物種の分類情報，Wikipedia における「記事概要」，および，Wikipedia 内でリンクが張られている「関連記事の項目名」である（表 2）．これらの取得できた情報を各生物の専門家に提示したところ，「記事概要」については数百文字程度の短い説明にとどまっていることなどから「情報が不十分である」との意見を得た．一方，「関連記事の項目名」については，「そのままでは，様々な関連語が列挙されているだけで不十分であるが，“形態”，“生態”など適切なカテゴリに分類されれば有効な情報となり得る」というコメントを得た．このことより，DBpedia をバイオミメティック・データベースにおいて利用するには，「関連記事の項目名」を適切なカテゴリに分類して用いることが重要であると言える．

また日本語 Wikipedia オントロジーは，DBpedia よりも詳細なプロパティ（他のデータとの関係）が定義されていることが特徴であり，今回対象とした生物種の情報では「近縁種」，「色」など DBpedia には定義されていないプロパティが複数見られた．また，各データ間の is-a 階層に関する情報（図 1 では jwo:hyper で表される）は，日本語 Wikipedia オントロジーの方が DBpedia よりも整備されているため，is-a 関係の判定にも，日本語 Wikipedia オントロジーが有用と思われる．

3.3 Linked Data 利用に向けた技術整備

前節で述べたような，既存の Linked Data の利用を検討するに辺り，Linked Data 技術の利用を補助するソフトウェアを開発した．開発した主なソフトウェアは「簡易 SPARQL ツール」および「簡易 LOD 検索サイト作成ツール」である．

「簡易 SPARQL ツール」は，任意の SPARQL エンドポイント（Linked Data を検索するための API）に対して，「キーワードによる Linked Data の簡易検索」ができるツール（Java によるクライアントアプリケーション）であり，検索にヒットしたデータ（Linked Data ではリソースと呼ばれる）一覧を選択することで，そのリソースの持つプロパティ一覧が表示されリンクを辿ることができ，「LOD の簡易ブラウザ」としても利用できる（図 2）．

その他にも，

－複数の SPARQL エンドポイントに対する横断検索

－あらかじめ用意してキーワードリストに対する一括検索

機能があり，前節で述べたような，複数の Linked Data に含まれるデータの一括検索が容易に行える．なお，前節の予備的調査は，本ツールを用いて行った．

一方，「簡易 LOD 検索サイト作成ツール」は，任意の SPARQL エンドポイントを対象とした「Linked Data の簡単な検索サイト」を，簡単なテンプレートに沿った Java Script の記述で作成できるツールである．サポートしている検索機能は，

－キーワード一致による検索

－選択した属性の組み合わせに一致するデータの検索（ファセット検索）

であり，簡単な設定ファイルを修正するのみで，Linked Data を利用した簡単なカタログサイトが作成できる．技術的には，サーブレットと Java Script の組み合わせにより実装されており，Linked Data の検索サイトを Java Script ベースで容易に開発できる．本ツールで作成したサイトの例[1]も公開しており，DBpedia を対象に生物種の情報に限定した検索を行うサービスのプロタイプ[2]も開発されている（図3）．バイオミメティック・データベースの開発は，本ツールを用いて実装される予定である．

なお，こられのツールはオープンソースソフトウェアとして，

http://sourceforge.jp/projects/easylod/

にて，公開されている．

4. まとめ

本稿では，工学研究者が生物多様性情報をもとに技術改革の着想を得ることを支援するバイオミメティック・データベースの開発について述べた．本データベースの基本的な考え方は，工学と生物学の双方の知識を領域横断的に体系化したバイオミメティック・オントロジーを利用者の視点に応じて探索して，得られた概念間のつながりを用いた検索を行う点にある．さらに，既存の Linked Data から生物に関連する情報を取得することで，オントロジーのみではカバーすることができない広範囲な領域の知識を利用することができる．

現状では，小規模なオントロジーを用いた様々な観点からの探索の試行と，既存の Linked Data のうち生物に関する概要情報が含まれる，Lodac Speices，DBpedia（日本語/英語）および日本語 Wikipedia オントロジーを対象に，バイオミメティク・デー

[1] http://lodosaka.hozo.jp/EasyLOD/

[2] http://lod.hozo.jp/SpeciesFinder/

タベース構築に有用な情報が取得可能であることの予備的検証を実施した. その結果, 先行研究[古崎 13b]で検討した, バイオミメティック研究者が必要とする検索が, これらの情報を適切に組み合わせることで実現可能であろうことが確認できた.

また, バイオミメティック・データベース構築に必要となる技術的整備として, Linked Data を用いたシステム開発に用いる基盤ソフトウェアを開発した. これらのソフトウェアは, データベースの試作に用いられると共に, バイオミメティクス以外の領域における Linked Data 技術を用いたシステム開発での利用が期待される.

今後の課題としては, 第一に, バイオミメティック・オントロジーの大規模化とそれに伴う既存 Linked Data の利用形態の検討が必要となる. 3 章で検討した各 Linked Data から取得できる生物に関する情報は, バイオミメティック・オントロジーの拡充に用いるという方法と, オントロジーと連携(マッピング)して利用するという方法の 2 通りが考えられる. その際には, バイオミメティック研究に特化した情報はオントロジーに取り込み, 生物一般に関する情報は既存 Linked Data とのマッピングで扱う, などの設計が必要となる.

第二には, 拡充したオントロジーと Linked Data を用いた探索・検索システムの開発を行う. 基本的な技術は既に開発済みであるが, 探索対象の大規模化に伴う探索範囲・方法の制御や, ユーザが直感的に利用できるインタフェースの設計が重要になると思われる.

さらに第三には, 同一プロジェクトで開発されているバイオミメティクスの画像データベースをはじめ, 文献, 標本などの外部データベースとメタデータを介した連携の仕組みの設計・開発を行う. 基本的な仕組みとしては, Linked Data を含む Semantic Web 技術の標準仕様に沿ったメタデータ付与を, 本研究で構築するバイオミメティク・オントロジーで定義された語彙を用いて行いことで柔軟な連携が行えると考えている.

そして, これらの一連のシステムを統合することでバイオミメティック・データベースを構築し, 利用者のフィードバックを受けつつ, 実用的なプラットフォームを実現することが本研究の最終的な課題となる.

謝辞

本研究の一部は科学研究費補助金 新学術領域研究(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」および, 基盤研究(B)25280081「オントロジーの多次元的視点管理に基づく領域横断型セマンティックデータの知的探索」の助成による.

参考文献

[大澤 14] 大澤剛士, 神保 宇嗣:ビッグデータ時代の環境科学ー生物多様性分野におけるデータベース統合, 横断利用の現状と課題ー, 数理統計, Vol.61, No.2, pp.217－231, 2013.

[下村 10] 下村政嗣:生物の多様性に学ぶ新世代 バイオミメティック材料技術の新潮流, 科学技術動向 Vol.110, pp.9-28, 2010.

[Kozaki 11] 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.

[古崎 13a] 北河祐作, 古崎晃司:大規模オントロジーの知的探索に向けた多段階展開型概念検索システムの開発, 人工知能学会研究会資料, SIG-SWO-A1203-09, 2013.

[古崎 13b] 古崎晃司, 他: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ック・オントロジーの試作, 2013 年度人工知能学会全国大会,3I1-3, 2013.

[多田 14] 多田恭平, 古崎晃司, 他:専門文書と Linked Open Data を用いたバイオミメティクス・オントロジーの大規模化の試み, 2014 年度人工知能学会全国大会, 2F1-5, 2013.

[玉川 11] 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, No.4, pp.504-517, 2011.

[ヒース 13] トム ヒース (著), クリスチャン バイツァー (著), 武田英明 (監訳):Linked Data: Web をグローバルなデータ空間にする仕組み, 近代科学社,2013

[南 11] 南佳孝, 加藤文彦, 大向一輝, 武田英明, 新井紀子, 神保宇嗣, 伊藤元己, 小林悟志:生物情報基盤構築に向けた生物関連データの Linked Data 化の取り組み, 第 26 回セマンティックウェブとオントロジー研究会, 人工知能学会, 2011.