

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 15 日現在

機関番号：14401
研究種目：基盤研究(B) (一般)
研究期間：2013～2016
課題番号：25280081
研究課題名(和文) オントロジーの多次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

研究課題名(英文) Intelligent Exploration of Semantic Data through Domains based on Multi-dimension Viewpoint Management of Ontologies

研究代表者
古崎 晃司 (Kozaki, Kouji)
大阪大学・産業科学研究所・准教授

研究者番号：00362624
交付決定額(研究期間全体)：(直接経費) 12,600,000円

研究成果の概要(和文)：本研究では、様々な対象領域の知識を機械可読な形式で格納した構造化データ(セマンティックデータ)を、利用者の目的や観点に応じて柔軟に、かつ、領域横断的に探索し、有用な知識を効率的に導き出すことを可能とする「オントロジーの多次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」技術を開発した。
これにより、様々なセマンティックデータを対象に、ユーザの関心に応じた視点から「知識のつながり」を見つけ出したり、比較したりすることが可能となり、様々な知的システム開発の基盤となるソフトウェアモジュールとして利用できるようになった。

研究成果の概要(英文)：In this research, we developed techniques for intelligent exploration of semantic data which is structured data in machine readable formats. The techniques enable the users to explore semantic data through domains based on multi-dimension viewpoint management of ontologies according to their interests.
As the result, the users can find and/or compare "a series of relationships between knowledge (chains of knowledge)" from various semantic data. It could be used as software modules which support bases of many intelligent systems.

研究分野：オントロジー工学

キーワード：オントロジー 知的探索 セマンティックデータ Linked Data 視点管理

1. 研究開始当初の背景

急激な勢いでコンピュータに蓄積され続ける情報を有益な「知識」として活用する技術の確立は、情報科学分野のみならず社会全体での急務である。特に、公共的なデータを公開し誰でも自由に使えるようにするというオープンデータの取り組みは、欧米を中心に急速に広がっており、国内でも、東日本大震災以降、総務省や経済産業省などでオープンデータに関する取り組みが盛んに進められている。

技術的には Web の創始者である Tim Barres-Lee が提唱した、Web 上で Open にした構造化データを相互に Link することにより様々な知識処理を可能とする Linked Open Data (LOD)の活用が急速に広まっている。

これらのデータを真に価値のある「知識」として活用するためには、データの意味(セマンティクス)を適切に与え、相互の意味的なつながりを明確にする必要がある。特に、複数の異なる領域のオープンデータを統合して利用するには、特定領域において膨大に収集されたデータの分析を効率的に行う量的問題のみならず、複数領域を横断した情報の意味的なつながりの活用を目指した質的問題の解決が強く求められている。

2. 研究の目的

本研究は、様々な対象領域の知識を機械可読な形式で格納した構造化データ(セマンティックデータ)を、利用者の目的や観点に応じて柔軟に、かつ、領域横断的に探索し、有用な知識を効率的に導き出すことを可能とする「オントロジーの多次元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」技術の確立を目的とする。

具体的には下記の項目について研究する。

- (1)セマンティックデータの知的探索の概念定義を行い、一般的な情報検索、推薦、可視化などの関連技術との概念的な相違点・関連性を明確にする。
- (2)オントロジー探索の結果を比較・集約する技術とその応用システムを開発する。
- (3)知的探索技術の対象を Linked Data 技術に拡張する。
- (4)知的探索技術を様々な形式の Open Data を適用するための技術的課題と解決法を明らかにする。
- (5)上記の成果に基づき知的探索システムを開発すると共に、具体的な専門領域(サステナビリティ学、等)へ適用し、システムの評価と洗練を行う。

3. 研究の方法

本研究は、研究代表者らがこれまでの進めてきた2つの技術を発展させることで行う。

1つ目の技術は、オントロジー工学の基礎理論に基づいて開発されたオントロジー構築ツール「法造」である。「法造」は対象世

オントロジー工学に基づく知識の構造化モデル

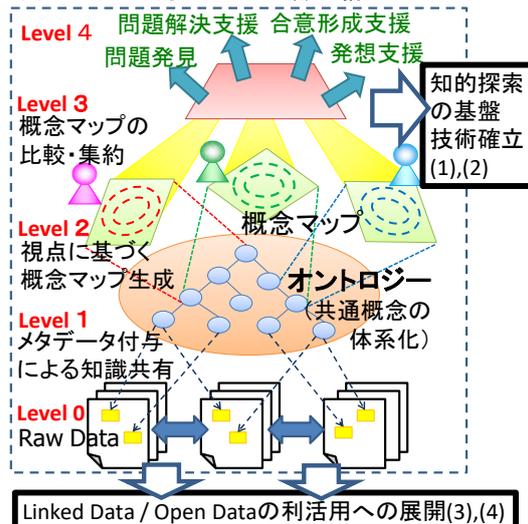


図1 オントロジー工学に基づく知識構造化モデルにおける本研究の位置づけ

界を捉える観点の違いを適切に表すことができるロール理論に基づいている点が特徴であり、国内唯一の本格的なオントロジー構築ツールとして国内外に約5,000の登録ユーザを有する。

2つ目の技術は、領域を横断した知識を様々な視点から構造化する手法を5つのレベルに分けてモデル化した「オントロジー工学に基づく知識の構造化モデル」(図1)である。このモデルは、多様な領域を横断した学際的な研究分野であるサステナビリティ・サイエンスを対象としたオントロジー構築研究成果として提案したものである。本研究開始時点において、視点に基づく概念マップ生成(Level.2)が「法造」の「オントロジー探索ツール」として実装されている。

本研究では、これら研究成果を発展させて、本研究の目的である複数領域を横断したセマンティックデータの知的探索の技術的を開発する。図1に示すように、研究目的の(1)および(2)は、知識の構造化モデルのLevel.3および4に相当する技術を開発すること相当し、研究目的(3)および(4)は、本技術の適用対象を Linked Data/Open Data へと展開することに相当する。

具体的には、下記の項目に沿って研究を進める。

- (1)セマンティックデータの知的探索の概念的・技術的な意味を明らかにする。知的探索を評価するための要素として一般的に、計算機モデル(Computational models)、人的要因(Human factors)、対象領域におけるニーズが重要とされる。また、知的探索と従来の検索の違いは、目的が事前に明確になっているか否か、システムからユーザへの提案の有無などから議論できる。

- (2)知識の構造化モデルのLevel.3および4

に相当する、「視点に基づくオントロジー探索結果（概念マップ）の比較・集約」手法の設計・開発を行い、問題解決支援や発送支援につなげる。その際には、これまでのユーザ試用実験において課題として得られた、オントロジー探索ツールのユーザインタフェースの抜本的な改良の必要性にも対応する。

- (3) 知的探索技術の適用対象をこれまでの「オントロジー（クラスレベルの知識）」から、Linked Data を中心とした「インスタンス間の関係（個々の具体的レベルの知識やデータ）」に拡張する。具体的には、
 - ・ Linked Data で定義された関係（リンク）を探索し、マップを生成する機能
 - ・ 必要に応じて、探索対象をクラスもしくは、インスタンスに限定できる機能を開発する。
- (4) 既存の Open Data を探索対象として利用できるようにするため、様々な Open Data を Linked Data に変換するシステムを開発する。これにより、項目③で開発した機能での知的探索手法の適用が可能となる。
- (5) 開発した知的探索システムを、これまで研究対象としてきたサステナビリティ・サイエンス分野を含め、いくつかの領域に適用して評価を行う。

4. 研究成果

- (1) セマンティックデータの知的探索の概念的・技術的な意味について考察した。

計算機モデル(Computational models)の観点からは、クラス（概念）レベル・インスタンスレベルの探索と、各レベルにおける探索方法の違いについて考察し、特に、クラスレベルにおいてはオントロジー記述言語の仕様に依って、抽出すべき関係性を整理することの必要性を明らかにした。実装レベルにおいては、オントロジーレベルの API や推論機構、RDF ベースのクエリ言語である SPARQL を利用する方法のいずれかを選択、組み合わせるという方針を決定した。

人的要因(Human factors)からは、オントロジーやセマンティックデータについての習熟度、対象領域における事前知識、関心がある概念の違いを検討する必要があると考えた。

また国際ワークショップ内で「視点に基づくオントロジー探索」手法について、本研究で開発したツールを利用したハンズオンセッションを開催し、利用者からのフィードバックを受けた。

これらを通して得られた知見は、書籍の一部として出版（図書①）するとともに、

これ以降に述べる研究成果へ反映されている。

- (2) オントロジー探索の視点を“注目する概念定義の構造を段階的に展開することで、探索範囲の意味的境界を適切に管理する技術”を利用し、視点管理・選択を円滑に行うことを可能とする「オントロジーに基づく多次的視点管理」手法を開発した。

さらに開発した「オントロジーに基づく多次的視点管理」手法、および「視点に基づくオントロジー探索結果（概念マップ）の比較・集約」手法を、利用者のフィードバックを受けて改良した。

特に領域の専門家からは、「探索結果である概念マップ全体を詳細に見るのではなく、その領域で注目している概念を中心に閲覧することで、全体像を捉えたい」という要望が強かった。そこで、探索結果から利用者が注目している概念のみをマップに表示し、詳細な関係は選択時のみに表示する機能を追加した。

最終的なツールは「法造」の拡張機能として実装し、フリーソフトウェアとして公開した。

- (3) 本研究における知的探索技術の適用対象をこれまでの「オントロジー（クラスレベルの知識）」から、Linked Data を中心とした「インスタンス間の関係（個々の具体的レベルの知識やデータ）」に拡張した。具体的には、

- ・ Linked Data への検索に用いる標準 API である SPARQL エンドポイントを利用した探索機能の Web 版（図 2）

- ・ 知的探索システムの機能を拡張した「オントロジーにおけるクラスレベルの知識」に加えて「具体事例に直結するインスタンスレベルの知識」を対象とした探索機能、および、特定の事例を中心とした探索範囲の適切な絞り込みが機能を開発した。特に Web 版にでは、探索結果



図 2 知的探索ツールの Web 版

を表示するだけでなく、外部 DB での検索結果など、概念マップを起点として、関連情報を表示する機能が求められ実装した。

(4) 提案手法の対象を、Linked Data に準拠した形式 (RDF) 以外の様々な形式で公開されている Open Data に拡張するため、Open Data の公開フォーマットとして利用されることが多い CSV 形式のデータを RDF 形式に変換するツール「CSV2LOD」を開発した。

本ツールの特徴は、セマンティックデータとしての構造化に重要な、標準的な語彙の設定を補助する推薦機能を有している点と、変換に用いる情報をテンプレートとして保存することで、元データが更新された際に、簡単に RDF 形式のデータも更新できる点にある。これにより、既存の Open Data を提案システムの探索対象として取り込むことが可能になった。

本ツールを利用したワークショップも数回開催し、既存の Open Data を Linked Open Data として公開する技術の普及にも貢献した。

(5) 完成したシステムを、昨年度までに試験的に適用したサステナビリティ学領域の応用事例に再適用することで、初期版で見られた問題点が改善されたことを確認した。

さらに、他領域での応用事例として、

- ・バイオミメティクス (生物を模倣した新技術開発) 分野における新規材料開発の発想支援 (雑誌論文③, 学会発表③)
- ・疾患オントロジーにおける異常状態の is-a 階層の閲覧, 疾患に現れる異常状態の因果連鎖の探索 (雑誌論文①, ④, 学会発表②)
- ・自治体が公開している広報情報のオープンデータの統合と分析 (学会発表①) に適用し、プロトタイプシステムを開発・公開した。

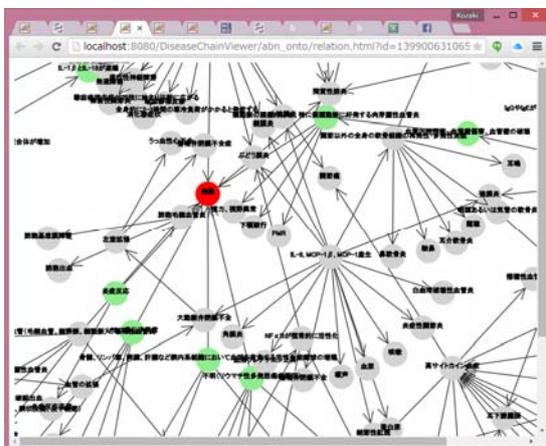


図 3 異常状態の因果連鎖の探索例

5. 主な発表論文等

[雑誌論文] (計 5 件)

- ① Kouji Kozaki, Yuki Yamagata, Riichiro Mizoguchi, Takeshi Imai, Kazuhiko Ohe, Disease Compass- A Navigation System for Disease Knowledge based on Ontology and Linked Data Techniques, 査読有, J. of Biomed Sem, 印刷中.
- ② 増田壮志, 古崎晃司, 概念階層の比較に基づくオントロジー品質向上支援手法, 人工知能学会論文誌, 査読有, 32(2), 2017, p.E-G71_1-10.
<http://doi.org/10.1527/tjsai.E-G71>
- ③ 古崎晃司, 來村徳信, 溝口理一郎, 生物規範工学オントロジーと Linked Data に基づくキーワード探索, 人工知能学会論文誌, 査読有, 31(1), 2016, p.LOD-D_1-12.
<http://doi.org/10.1527/tjsai.LOD-214>
- ④ 山縣友紀, 古崎晃司, 今井健, 大江和彦, 溝口理一郎, 疾患知識統合に向けた異常状態オントロジーの Linked Data 化, 人工知能学会論文誌, 査読有, 31(1), 2016, p.LOD-A_1-15.
http://doi.org/10.1527/tjsai.31-1_LOD-A
- ⑤ Riichiro Mizoguchi, Antony Galton, Yoshinobu Kitamura, Kouji Kozaki, Families of roles: A new theory of occurrent-dependent roles, Applied Ontology, 10(3-4), 2015, pp.367-399.
<http://doi.org/10.3233/A0-150148>

[学会発表] (計 49 件)

- ① 古崎晃司, 地域課題分析に向けた自治体広報情報の収集と LOD 化の試み, 第 38 回セマンティックウェブとオントロジー研究会, 宮古島 (沖縄県), 2016 年 2 月 18 日-19 日.
- ② Kouji Kozaki, Disease Compass - a navigation system for disease knowledge based on ontology and linked data techniques, The 6th International Conference on Biomedical Ontology (ICBO2015), 5 pages, Lisbon, Portugal, July 27-30, 2015.
- ③ Kouji Kozaki, A Keyword Exploration for Retrieval from Biomimetics Databases, The 4th Joint International Semantic Technology (JIST2014), Chiang Mai, Thailand, Nov. 9-11, 2014.
- ④ Kouji Kozaki, Choosing Related Concepts for Intelligent Exploration, The 4th Joint International Semantic Technology (JIST2014), Chiang Mai, Thailand, Nov. 9-11, 2014.
- ⑤ Kouji Kozaki, Ontology Engineering for Big Data, Ontology and Semantic Web for Big Data (ONSD2013) Workshop, Bangkok, Thailand, Sep. 5, 2013.

[図書] (計 2 件)

- ① Jenny S. Huang, Kouji Kozaki, Terukazu Kumazawa (Eds.:Francisco J. Garcia-Penalvo, Alicia Garcia-Holgado), IGI Global, Open Source Solutions for Knowledge Management and Technological Ecosystems, Chapter 8 Knowledge Structuring for Sustainable Development and the Hozo Tool, 2017, pp.195-221.

[その他]

オントロジー構築利用環境「法造」の公開サイト, <http://www.hozo.jp>
本研究課題で開発した LOD 構築用ツールの公開サイト, <http://lod.tech>,
<https://ja.osdn.net/projects/easylod/>

6. 研究組織

(1) 研究代表者

古崎 晃司 (KOZAKI, Kouji)
大阪大学・産業科学研究所・准教授
研究者番号：00362624

(2) 研究分担者

来村 徳信 (KITAMURA, Yoshinobu)
立命館大学・情報理工学部・教授
研究者番号：20252710

(3) 研究分担者

熊澤 輝一 (KUMAZAWA, Terukazu)
総合地球環境学研究所・研究基盤国際センター・准教授
研究者番号：90464239

(3) 連携研究者

山本 泰智 (YAMAMOTO, Yasunori)
大学共同利用機関法人情報・システム研究機構 (機構本部施設等)・データサイエンス共同利用基盤施設・特任准教授
研究者番号：50470076
(平成 27 年度より連携研究者)