

平成 28 年 6 月 13 日現在

機関番号：62615  
研究種目：基盤研究(C) (一般)  
研究期間：2013～2015  
課題番号：25330369  
研究課題名(和文) Linked Data間の意味的關係付け手法に関する研究

研究課題名(英文) Semantic Interlinking on Linked Data

研究代表者

市瀬 龍太郎 (Ichise, Ryutaro)

国立情報学研究所・情報学プリンシプル研究系・准教授

研究者番号：00332156

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：Linked Data は、データのWeb と呼ばれ、急速に普及しつつあるが、データ間のリンク(関係付け)が不十分なため、実用上で大きな問題となっている。本研究では、大規模で様々な種類のデータが分散して提供されているLinked Data に対して、意味的な関係付けを高速かつ高精度に行う手法を開発することが目的である。そのために必要な新たな技術の開発を行った。

研究成果の概要(英文)：The number of entities on Linked Data is rapidly growing recently. In order to use Linked Data practically, we need to have enough links between entities. We developed novel methods for interlinking entities on Linked Data.

研究分野：人工知能

キーワード：人工知能 セマンティックWeb Linked Data

## 1. 研究開始当初の背景

Web では、様々な情報が配信され、それらの情報が企業や人々の意思決定に大きな役割を果たしている。しかし、現在の Web では、意思決定をする際に、必要な情報を人手で収集し、それらを統合、分析しなければならない。そのような作業を自動化するために、セマンティック Web では、様々なデータを Linked Data と呼ばれる形式で公開している。Linked Data は、地理情報、人や企業の情報、遺伝子やたんぱく質の情報、世論調査結果など、様々なデータを特定のデータ形式で表したものであり、これらをお互いに関係付けることで、意味的な処理を機械に行わせる基盤としている。この試みは 2007 年頃から始まり、2011 年秋の時点では、316 億ものデータが公開され、急速に広まっている。例えば、医療に関しての Linked Data を利用すると、従来別々のデータベースで取り扱われていた臨床試験、薬、病気のデータを組み合わせることができ、医薬品の副作用に関する情報の分析が容易になる。しかし、このようなデータは、それぞれ違う組織や人が作成しているため、データの記述に関して相違が生まれる。上記の例では、同じアルツハイマー病に関して、「Alzheimer 's disease」と記述されたり、「Alzheimer\_disease」と記述されたりする。Linked Data では、データ同士の関係を同定し、関係に関する記述を与えることで、機械的な意味処理を可能としている。そのために、従来は、人手でデータ同士を関係付けたり、関係付けを行う規則を人手で作成したりすることで対応していた。しかし、近年のデータの多様化、急速な増大に伴い、このような人手を介する対応は、困難になってきている。そこで、Linked Data 同士の意味的な関係を同定し、自動的にデータ同士を関係付ける手法の開発が喫緊の課題となってきた。

## 2. 研究の目的

Linked Data は、データの Web と呼ばれ、急速に普及しつつあるが、データ間のリンク（関係付け）が不十分なため、実用上で大きな問題となっている。本研究では、大規模で様々な種類のデータが分散して提供されている Linked Data に対して、意味的な関係付けを高速かつ高精度に行う手法を開発することが目的である。そのために、オントロジー・アライメントで使われる機械学習技術とインスタンス・マッチングで使われるブロッキング技術を融合することで新たな意味的な関係付け技術を開発する。

## 3. 研究の方法

3 年間の研究は、下記のような形式で遂行した。

### (1) 2013 年度

2013 年度は、研究の開始年に当たるため、主に研究に必要な環境の整備

に焦点を当て、下記の 2 つに分けて研究開発を実施した。

### Linked Data 分析基盤の構築

本研究では、インターネット上で分散して公開されている Linked Data を研究の対象としている。そのため、最初にデータを収集し、研究用の環境の構築を行う。研究に使うためには、収集したデータを適切なフォーマットに変換したり、不要な部分を除去したりするなどの前処理が必要となる。そのような作業を施した上で、研究用の Linked Data のデータベースを構築した。さらに、構築したデータベースを利用して、データの特性を分析するために、類似度を視覚化するシステムや検索システムの作成を行った。単体技術による性能評価とその拡張手法の開発

オントロジー・アライメント手法を適用し、2 つのデータの関係の発見を試みた。その過程だけでは、直接の対応関係が見つからない場合に、近傍を効果的に探索することで、新たな関係を発見するアルゴリズムの開発を行った。

### (2) 2014 年度

2014 年度は、2013 年度に整備した研究環境を利用して、以下の 2 つに分けて研究開発を行った。

大規模データに対応するための研究基盤の拡張と高度化

本研究で用いる Linked Data は急速に普及しており、取り扱うデータ量が拡大し続けている。研究を効果的に推進するためには、大規模データを効率的に取り扱う研究基盤が欠かせない。そのため、2013 年度に構築した検索システムを拡張し、効率を改善する機構の作成を行った。

オントロジー・アライメント手法とインスタンス・マッチング手法の融合化

2 つの手法を新たなアプローチで融合することにより、より精度の高い意味的な関連付け技術の開発を行い、様々な基礎的データを得た。

### (3) 2015 年度

2015 年度は、研究の最終年度に当たるため、これまでの研究成果を利用しながら、以下の 2 つに分けて研究開発を行った。

取り扱いデータの大規模化対応機構の研究

大規模な Linked Data を取り扱うためには、時間計算量、空間計算量の双方を削減する必要がある。そのための機構の開発に取り組んだ。具体的には、これまでに研究してきたシステムの一部のアルゴリズムを改善することにより、両方の計算量を大きく減らしたシステムの作成を行った。精度向上のための機械学習機構の研究

これまでに開発してきたシステムの精度をさらに高めるための手法について、研究を行った。意味的な関係付けを行う際に、システムは様々な設定を利用する。その設定状況により、システムの性能が変化するため、自動的に適切な設定を行うことができる学習アルゴリズムの開発を行った。その結果、システムが精度高く関連付けを行うことが可能となった。

#### 4. 研究成果

##### (1) 2013 年度

前章で述べた研究で得られた成果に基づき、論文 3 本の出版を行った。

##### (2) 2014 年度

前章で述べた研究で得られた成果に基づき、論文 2 本の出版を行った。

##### (3) 2015 年度

前章で述べた研究で得られた成果に基づき、論文 2 本の出版と 2 件の学会発表を行った。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 7 件)

1. Khai Nguyen, Ryutaro Ichise: Linked Data Entity Resolution System Enhanced by Configuration Learning Algorithm, IEICE Transactions on Information and Systems, Vol. E99-D, No. 6, 2016, (査読有, 採録決定済, 出版予定)
2. Khai Nguyen, Ryutaro Ichise: Heuristic-based Configuration Learning for Linked Data Instance Matching, Proceedings of the 5th Joint International Semantic Technology Conference, pp.56-72, LNCS 9544, Springer, 2015, (査読有)
3. Md-Mizanur Rahoman, Ryutaro Ichise: inteSearch: An Intelligent Linked

Data Information Access Framework, Proceedings of the 4th Joint International Semantic Technology Conference, pp.151-163, LNCS 8943, Springer, 2014, (査読有)

4. Md-Mizanur Rahoman, Ryutaro Ichise: Automatic Inclusion of Semantics over Keyword-based Linked Data Retrieval, IEICE Transactions on Information and Systems, Vol.E97-D, No.11, pp.2852-2862, 2014, (査読有)
5. Md-Mizanur Rahoman, Ryutaro Ichise: TLDRet: A Temporal Semantic Facilitated Linked Data Retrieval Framework, Proceedings of the 3rd Joint International Semantic Technology Conference, pp.228-243, LNCS 8388, 2013, (査読有)
6. Natthawut Kertkeidkachorn, Ryutaro Ichise, Atiwong Suchato, Proadpran Punyabukkana: An Automatic Instance Expansion Framework for Mapping Instances to Linked Data Resources, Proceedings of the 3rd Joint International Semantic Technology Conference, pp.380-395, LNCS 8388, 2013, (査読有)
7. Vaclav Jirkovsky, Ryutaro Ichise: MAPSOM: User Involvement in Ontology Matching, Proceedings of the 3rd Joint International Semantic Technology Conference, pp.348-363, LNCS 8388, 2013, (査読有)

[学会発表](計 2 件)

1. Khai Nguyen, Ryutaro Ichise, ScSLINT: Time and Memory Efficient Interlinking Framework for Linked Data, Proceedings of the ISWC 2015 Poster and Demonstrations Track, 2015, (USA, Bethlehem), 2015/10/11 - 2015/10/15, (査読有)
2. Khai Nguyen, Ryutaro Ichise: A Heuristic Approach for Configuration Learning of Supervised Instance Matching, Proceedings of the ISWC 2015 Poster and Demonstrations Track, 2015, (USA, Bethlehem), 2015/10/11 - 2015/10/15, (査読有)

[図書](計 0 件)

[産業財産権]

出願状況(計 0 件)

名称:  
発明者:  
権利者:  
種類:

番号：  
出願年月日：  
国内外の別：

取得状況（計 0 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

#### 6. 研究組織

##### (1) 研究代表者

市瀬 龍太郎 (ICHISE Ryutaro)  
国立情報学研究所・情報学プリンシプル研  
究系・准教授  
研究者番号：00332156

##### (2) 研究分担者

該当なし

##### (3) 連携研究者

該当なし