

氏 名 LIHUA ZHAO

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 1636 号

学位授与の日付 平成25年9月27日

学位授与の要件 複合科学研究科 情報学専攻  
学位規則第6条第1項該当

学位論文題目 Ontology Integration for the Linked Open Data

論文審査委員 主 査 准教授 市瀬 龍太郎  
教 授 武田 英明  
准教授 大向 一輝  
教 授 相澤 彰子 国立情報学研究所  
准教授 古崎 晃司 大阪大学

論文内容の要旨  
Summary of thesis contents

The Linked Open Data (LOD) cloud contains tremendous amounts of interlinked instances, from where we can retrieve abundant knowledge. In order to access to the linked data, we have to be familiar with the ontology of each data set. However, because of the heterogeneous and big ontologies, it is time consuming to learn all the ontologies manually and it is difficult to observe which properties are important for describing instances of a specific class. In order to construct an ontology that can help users easily access to various data sets, we propose the Framework for InTegrating ONtologies (FITON) that can reduce the heterogeneity of the ontologies, retrieve core ontology schemas, and construct easily understandable integrated ontology. The integrated ontology constructed by FITON can help us discover missing links, detect misused properties, recommend standard ontology schemas for the instances, and improve the information retrieval with simple SPARQL queries.

The thesis consists of 6 Chapters. In the following, we introduce each Chapter.

- In Chapter 1, we introduce background knowledge about the Linked Open Data, OWL, and SPARQL. Then we present the motivation, the problem statement, and contributions of this thesis. Our research mainly focuses on solving three problems: ontology heterogeneity problem, difficulty in identifying core ontology schemas, and missing domain or range information problem. The three main components of FITON solve each problem, which are graph-based ontology integration, machine-learning-based approach, and integrated ontology constructor.
- In Chapter 2, we introduce some ontology matching methods that are commonly used in solving ontology heterogeneity problem. The ontology matching methods include string-based ontology similarity measures and WordNet-based (knowledge based) ontology matching methods. Then we discuss some research work about analyzing the SameAs networks that can assist finding ontology alignments from interlinked instances and discover SameAs links between instances. We also introduce some approaches which extract concepts from data or ontologies using machine learning methods. At last, we list some unsolved problems in the related work that can be solved with our approach and summarize this chapter.
- In Chapter 3, we present two approaches that can solve the ontology heterogeneity problem in the LOD cloud by integrating the heterogeneous ontologies. We introduce

the Mid-Ontology learning approach and graph-based ontology integration approach that solve the ontology heterogeneity problem. The Mid-Ontology Learning approach finds alignments between different ontologies in the LOD data sets. This approach collects interlinked instances and applies ontology similarity matching methods on the collected triples to find related ontology predicates and integrates them into one ontology. The graph-based ontology integration method overcomes some limitations of the Mid-Ontology learning approach by performing ontology matching methods on the SameAs graph patterns. This method integrates both related classes and properties, and does not require a hub data set to collect data. Furthermore, different similarity measures are applied to the categorized <Predicate, Object> pairs and also default domain information is added to the integrated ontology for better understanding. The comparison results with the Mid-Ontology learning approach are also discussed to show the advantages of the graph-based ontology integration approach.

- In Chapter 4, we introduce machine-learning-based approach and integrated ontology constructor that can enrich the integrated ontology automatically. Although the integrated ontology constructed using the graph-based ontology integration approach successfully integrates related ontology schemas from various data sets, the integrated ontology may miss some core ontology schemas that are important for describing instances of the data sets. Furthermore, because of missing domain or range information in the definitions of the ontology schemas, it is difficult to observe the relations between properties and classes or the ranges of property values. In order to solve these two problems: the difficulty in identifying core ontology schemas and missing domain or range information, we propose two approaches to enrich the integrated ontology. The former problem is solved using the machine-learning-based approach and the latter one is solved by the integrated ontology constructor. Detailed experimental results and discussion are also included in this chapter.

- In Chapter 5, we introduce the Framework for InTegrating ONtologies – FITON that consists of three main components: graph-based ontology integration, machine-learning-based approach, and integrated ontology constructor. By combining these three components, we can solve the three proposed main problems. We compare the ontology alignment results with other ontology matching tools and also evaluate the integrated ontology with manually created ontology reference alignments. At last, we discuss the performance on different types of the data sets, possible applications for link discovery, and the effectiveness in information retrieval with the integrated ontology using some SPARQL examples.

(別紙様式 2)  
(Separate Form 2)

- In Chapter 6, we discuss the advantages of our approach by comparing with related work and also discuss remaining problems. Furthermore, we present the contributions and limitations of our research work and propose future work to achieve our vision. In addition, we propose possible further research by using the integrated ontology constructed using FITON.

博士論文の審査結果の要旨

Summary of the results of the doctoral thesis screening

インターネットでは、さまざまなインスタンスに対するデータが、**Linked Open Data**として公開されている。それを利用することで、インスタンスに関する知識を得ることができる。しかし、**Linked Open Data**を利用するためには、それぞれのデータがどのように記述されているかを表す多様なオントロジーを理解しなければならない。博士論文では、ユーザが容易に**Linked Open Data**を利用できるようにするために、多様なオントロジーを統合するという問題に取り組んでいる。中核となるオントロジーを抽出し、オントロジーの多様性を減らした統合オントロジーを構築することで、データ記述に用いられているオントロジーの理解が容易になる。そのための効果的な枠組みを構築したというのが、本論文の主張である。

本論文は、全6章からなる。第1章「**Introduction**」では、本研究の背景、動機について説明し、この論文で取り組む研究課題を、(1) オントロジーの多様性問題、(2) 中核的なオントロジーの同定問題、(3) 定義域・値域の欠損問題、の3点に整理し、この論文の成果を明らかにしている。

第2章「**Related Work**」では、この論文で述べられる手法の理解に必要な関連研究について述べている。まず、オントロジーの多様性問題を解くのに使われるオントロジーマッチング技術について述べている。次に、**Linked Open Data**をグラフとして分析することで、オントロジー間の関係を抽出する技術について述べている。次に、機械学習を使ってオントロジーやデータから概念を取り出す研究について述べている。そして、それらの研究で解決されていない問題を述べ、博士論文の研究上の位置づけを明らかにしている。

第3章「**Integrating Heterogeneous Ontologies**」では、上述の(1)の課題に取り組んでいる。ここでは「**Mid-Ontology 学習アプローチ**」と「**グラフベースオントロジー統合アプローチ**」の2つのアプローチについて述べられている。「**Mid-Ontology 学習アプローチ**」では、リンクの張られたインスタンスを集め、類似度を測ることによって関連する述語を見つけ、一つのオントロジーに統合する手法を提案している。そして、実験的にその手法の有効性を確認している。「**グラフベースオントロジー統合アプローチ**」では、**Linked Open Data**のグラフを利用することで、クラスとプロパティの両方の統合を実現し、さらにハブデータを仮定しなくて動作するように改良している。そして、実験的にその手法の有効性を確認している。

第4章「**Enriching the Integrated Ontology**」では、上述の(2)、(3)の課題に取り組んでいる。まず、(2)に対処するために、機械学習を適用する手法を提案し、実験的にその手法の有効性を確認している。次に、(3)に対処するために、統合オントロジー構築機を提案し、実験的にその手法の有効性を確認している。

第5章「**Framework for InTegrating ONtologies - FITON**」では、第3章で述べた「**グラフベースオントロジー統合アプローチ**」の手法と第4章で述べた2つの手法を統合したオントロジー統合フレームワーク **FITON**を提案し、実験的に提案手法の有効性を確認している。

第6章「**Conclusion**」では、博士論文の結論をまとめている。さらに、今後の展望につ

(Separate Form 3)

いて述べている。

上記のように、本博士論文は、3つの手法を統合することによって、**Linked Open Data**において、従来研究よりも効果的にオントロジー統合を実現可能であることを示した点で、この研究分野の発展に貢献するものである。また、この研究は、今後、普及が見込まれる**Linked Data**の基盤となる研究となるため、基盤技術開発という観点からも意義があると認められる。さらに、博士論文の内容は、1本の査読付ジャーナル論文、3本の査読付国際会議論文、2本の国内会議論文、2本の査読付ポスター論文として発表されており、社会からも評価されている。以上より、本論文は博士論文として、十分な水準であると審査委員全員一致で認められた。