

# メタデータの参照関係とスキーマに基づく LOD 間の類似性および 併用可能性算出手法

## Calculating similarity and combinability of LOD datasets based on their metadata schemas and link relationships

学籍番号 : 201821639

氏名 : 山中 勇樹

Yamanaka Yuki

Web 上でデータを公開し、共有するための仕組みとして様々な分野のデータを共通の形式で記述し、関連のあるデータ同士をリンクさせた Linked Open Data(LOD)が注目されている。LOD の仕組みに則って記述されたデータセット(LOD データセット)は Web アプリケーション開発や統計分析などに利用する場合、複数のデータセットを組み合わせることが容易である。そのため、Web 上に公開されている既存の LOD データセットを再利用し、LOD データセット同士やその他の形式の API などと自由に組み合わせて使用されている。

既存の LOD データセットは専用の検索サイトで検索することが可能であるが、併用可能な LOD データセットの例を提示しているサイトは少ない。これは、実際に組み合わせて使用された LOD データセットの例があまり公開されておらず、どのようなデータセットを併用することができるかが明確になっていないためである。そのため LOD の自由度の高さが十分に生かされていないのが現状である。この問題を解決するために、どのような LOD データセットが併用可能かを整理し、その指標を提案することが重要であると考えた。

本研究では併用可能な LOD データセット発見の支援を目的として、LOD データセットの構造を定義するメタデータスキーマや異なるデータセット間の参照関係に基づいた LOD データセット間の類似性および併用可能性を算出する手法を提案した。また、それに基づいて実際の LOD データセット間における類似性・併用可能性を算出する実験を行なった。結果、あらかじめ類似性や併用可能性が高いと予想される組の数値は他のものを上回った一方で、その数値は特定のパラメータに大きく依存する傾向があることを確認した。

研究指導教員 : 森嶋 厚行

副研究指導教員 : 永森 光晴