

# Linked Data クエリ構築支援のための日本語文に基づくグラフ構造の生成

## A Method of Generating Graph Structures from Japanese Sentences for Supporting of Building Queries of Retrieving Linked Data

学籍番号：201521648

氏名：YAN Jianan

Linked Data 応用システムの開発者はスキーマ定義が記述されていない Linked Data のデータセットに対して様々なパターンの検索を試行錯誤し、スキーマを推定しながらクエリを組み立てる。それは手間がかかるため、本研究ではその検索の手間を減らすことを目指す。本研究の目的は、Linked Data の構造を指定する検索クエリの書き方を知っている開発者向けに、日本語文からスキーマ定義が不明な Linked Data のデータセットを検索するためのクエリのグラフ構造を生成し、クエリの組み立てを容易にすることである。本研究では Linked Data のデータセットに RDF モデルを、クエリ言語に SPARQL を使用しているものを対象とする。

本研究では、日本語の情報要求から検索クエリのグラフを生成し、開発者がそのグラフに基づいて未知の述語を探し、クエリの作成を支援する。提案手法では、まず日本語の情報要求を構文解析し、係り受け関係を得る。各文節間の係り受け関係に基づいて、グラフを生成するルールを定める。次に、グラフ生成ルールを適用した結果を組み合わせて検索クエリのグラフを生成する。

Dbpedia 2015-4 のデータセットを情報要求に対する対象データとして、それに対して検索結果が得られると考えられる日本語文の情報要求を用いて、グラフ生成の評価実験を行った。まず、正解となる実験の比較対象を設定するため、グラフ生成ルールを使わずに人手で情報要求に基づいて対象データを検索するクエリを構築した。次に、提案したグラフ生成ルールに沿ってグラフを生成するプログラムを作成した。比較対象と同じ情報要求例を用いてそのプログラムでグラフを生成し、それに沿ってクエリを組み立て、検索を行った。そして、そこで得た検索結果数 A、正解の検索結果数 B、正解の結果と一致した数 C をそれぞれ求め、 $C/B$  を再現率、 $C/A$  を適合率とした。3 件の情報要求の中に再現率が 50%以上となったのは 2 件であり、適合率が 50%以上となったのは 1 件である。1 件についてはどちらも 100%となったので、その情報要求のパターン「名詞+を+動詞+名詞+の+名詞+を知りたい」に対して有効であることを確かめるため、同じパターン別の情報要求を用いて実験を行い、本手法がそのパターンの情報要求に対して有効であることを確かめた。一方、トリプル生成ルールの修正と追加についての検討が課題として残された。

研究指導教員：阪口 哲男

副研究指導教員：森嶋 厚行