

## Linked Data クエリ構築支援のためのグラフ構造生成系の開発

島田 慎太郎

計算機が解釈可能な形式でデータを流通・共有する仕組みの一つに、Linked Data がある。近年、Linked Data のデータセットが複数組み合わせられ、アプリケーション開発に利用される機会が増えてきた。本研究では、Linked Data のデータセットを利用するシステム開発者がその検索を行う場面に注目した。一般に、開発者はデータセット毎のスキーマ定義に従い、クエリを組み立てる。そこでは、Linked Data 検索に用いられるクエリ言語 SPARQL が用いられることが多い。しかし、スキーマ定義が記述されていない場合、開発者が様々なクエリを試行錯誤し、その検索結果からスキーマ定義を推定しなければならない。

先行研究では、開発者の手間を減らす手法として、情報要求を示す日本語文からグラフを生成してクエリの構築を支援するという提案がなされた。この提案手法では、グラフを生成するルールを手続きで表し、そこに情報要求文の係り受け解析結果を与え、グラフを生成している。そこでは、ルールの修正と追加に関する課題が残されていた。

本研究では、先行研究の手法に基づいた上で、更にルールの記述および追加を容易にし、多様な情報要求文に対応可能にすることを目的とする。その実現方法として、ルールによるグラフ構造の生成という手法を引き継ぎ、論理型言語によるグラフ構造生成系の開発を行った。まず、日本語の情報要求文の係り受け解析結果を、論理型言語の事実を記述する形式に変換するプログラムを作成した。次に、クエリ構築を支援するグラフを生成するルール記述を、第一階述語論理に基づいて設計・実装した。ルールを論理式で記述することで、手続きによる設計よりも、ルールの修正と追加が容易になることを目指した。

以上のようにして開発したグラフ構造生成系について、日本語による情報要求文を用いて、システムの評価実験を行った。情報要求文に基づいて対象データセットを検索するクエリを人手で構築して検索し、正解とした。次に、開発したグラフ構造生成系で同様の情報要求文を処理して、グラフを生成した。生成したグラフに沿ってクエリを組み立て、対象データセットの検索をした。その結果から得られた検索結果数を A、正解の検索結果数を B、正解と一致した検索結果数を C とし、 $C/B$  を再現率、 $C/A$  を適合率とした。

本研究の生成系では、新たに 5 種類のルールを追加した。結果から「名詞+を+動詞+名詞+を+動詞+名詞+の+名詞+を+知りたい」や「名詞+助動詞+名詞+助動詞+名詞+の+名詞+を+知りたい」というパターンが再現率と適合率が共に 100%となった。このように本研究で構築した生成系により、ルールを論理式で記述して、グラフ構造を得ることができるようになった。一方で、開発者が生成されたグラフからクエリを構築する際、適切な述語を選択させるための支援方法について課題が残った。

(指導教員 阪口哲男)