

メタデータ横断利用のための語彙マッピング方式と検索手法
- 異なる語彙で表現された歴史リソースの統合的利用環境の実現 - *

馬場裕子 (学生番号 200821671)

研究指導教員：杉本重雄

副研究指導教員：永森光晴

1. はじめに

Semantic Web の研究分野では、Linked Data が注目を集めている。Linked Data とは、RDF データを Web 上で公開し、関連のあるデータ同士を結びつけ、共有する活動のことである。そのような背景の中、様々な分野の機関で作成されたメタデータが、Web 上に公開され、利用できるようになることが予想される。しかしウェブ上にあるメタデータを横断的に用いるのは容易ではない。その理由は、機関ごとに用いているメタデータスキーマが異なるためである。メタデータスキーマが異なると、メタデータの語彙や構造に違いが生じてくるため、処理や検索のアルゴリズムが異なるという問題が起きる。

本研究では、異なるメタデータスキーマで書き表されたメタデータの語彙や構造の違いを吸収するための手法を提案し、メタデータの横断的な利用が行える環境の構築を行った。具体的には、同じ意味を持つ用語同士を機械が利用できるように、メタデータの語彙や構造をマッピングし、マッピングデータを用いて、メタデータを横断的な検索の手法を考案した。また、その手法を基礎として、メタデータ横断的検索システムの構築を行った。

2. メタデータ横断利用における問題点

まず、メタデータ横断利用における問題点の分析を行った。その結果、1)語彙の違い、2)メタデータの構造(要素の組み合わせ方)の違い、3)メタデータに記述された値の表記の違い、などの問題点があることがわかった。本研究ではそれぞれの問題点に対する解決方法として、1)同じ意味をもつ語彙のマッピング、2)メタデータを構成する要素の組み合わせ方を記述し、その中で同じ意味をもつ要素のマッピング、3)メタデータに記述された値の自動変換を行うことに決めた。

3. 語彙のマッピングフレームワーク

メタデータスキーマの語彙のマッピングは、メタデータ語彙のオントロジー構築フレームワークである VMF[1]を用いて行った。またこれに加えて、本研究ではメタデータの構造の違いの問題点を解決するためのマッピングフレームワーク SVT matrix の考案を行った。本研究で考案したマッピングフレームワークは、メタデータの要素の組み合わせ方を記述し、同じ意味をもつ組み合わせ同士で、等しい要素をマッピングするものである。表 1 に SVT matrix で定義されている主な要素を示し、図 1 に SVT matrix の記述例を示す。SVT matrix では、ある記述単位に対する、メタデータの要素の組み合わせ方を、`rdf:Statement` を用いた 1 つ以上

* "A Vocabulary Mapping Framework and Searching Methods for Metadata Crosswalking - Development of a System for Using Metadata about Historic Events in Multiple Metadata Schemas -" by Yuko BABA

表1 SVT matrix で定義されている用語一覧(一部)

要素名	意味
svt:Pattern	グラフパターンを表現するクラス
svt:Variable	グラフパターンにおいて、URI をもったノード、リテラルのいずれかにマッピングするリソースを表現するクラス
svt:hasPattern	グラフパターンを構成するトリプルの情報を参照するプロパティ

```

ex:pattern1 a svt:Pattern;
            svt:hasPattern [
                a rdf:Statement;
                rdf:object ex:var2;
                rdf:predicate ex:hasAuthor;
                rdf:subject ex:var1 ].
ex:var1 a bm:Variable .
ex:var2 a bm:Variable .
ex:pattern1 owl:sameAs ex:pattern2.
ex:var1 owl:sameAs ex:var3.
    
```

図1 SVT matrix の記述例 (一部)

のトリプルのパターンを用いて表現する。組み合わせが記述されたリソース間で、同じ意味を持つもののマッピングを行う。本研究ではこれらのフレームワークを用いて、10種類のスキーマを対象にマッピングデータを作成した。

4. クエリ生成手法

本研究で提案するクエリ生成手法では、任意のクエリから、それぞれのメタデータスキーマに対応したクエリを複数生成する。クエリ生成は次の手順で行う。(1)変換対象のクエリの構文解析を行い、クエリの検索条件に含まれる語彙(クラスやプロパティ)や、グラフパターンの情報を取得する。(2)取得した情報に関して、マッピングデータで表現された同義語、上位語、下位語の関係を類似語彙として検索する。また、変換対象のクエリのグラフパターンを示した svt:Pattern のインスタンスを調べ、それと同じ意味をもつインスタンスの情報を、クエリと類似した構造の情報として得る。(3)検索して得られた結果を、変換対象のクエリの語彙や構造と、置き換えたクエリを生成する。また本研究では

これに加えて、クエリに含まれている語彙に関する OWL の論理表現を用いて、推論を行い検索するクエリ手法の考案も併せて行った。

5. メタデータ横断利用システムの構築

本研究で提案した語彙マッピング方式と検索手法の実現例として、歴史に関するメタデータの統合的利用環境の実現をおこなった。歴史に関するメタデータスキーマには、一昨年度本研究室で開発した V-HAWKS[2]や、人物情報に関するスキーマである FOAF など、10種類のスキーマを対象としている。スキーマを分析し、3章に示したマッピング方式に従い、マッピングデータの作成を行った。また、それを基に、4章で示したクエリ生成手法に基づいて SPARQL クエリを複数生成し、歴史のメタデータを横断的に検索できるシステムの構築を行った。その結果、複数のスキーマに対応したクエリが生成され、より多く関連情報の結果が得られたことから、本研究で提案する手法で、メタデータ横断利用が行えることが確認できた。

6. まとめ

本研究で提案したマッピング方式と検索手法を用いて生成されたクエリにより、V-HAWKS や DBpedia など複数のコミュニティのメタデータから、横断的にデータを取得することが可能になった。今後の課題は、マッピングデータの作成支援や、用語の類似度を算出し、クエリの質を向上されること、などが考えられる。

文献

[1] CDLR. Vocabulary Mapping Framework (VMF) project. <<http://cdlr.strath.ac.uk/VMF/>>. (参照:2011-02-04)

[2] 馬場裕子, 永森光晴, 杉本重雄. RDF/OWL を利用した歴史情報閲覧システムの構築. 情報処理学会研究報告 人文科学とコンピュータ. 2008-CH-79, 2008年, pp.49-56.