

平成 26 年 6 月 19 日現在

機関番号：12102

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500295

研究課題名(和文) Linked Data 環境の実現に向けたメタデータスキーマレジストリの応用

研究課題名(英文) A Realization of the Linked Data Environment based on Metadata Schema Registries

研究代表者

永森 光晴 (Nagamori, Mitsuharu)

筑波大学・図書館情報メディア系・講師

研究者番号：60272209

交付決定額(研究期間全体)：(直接経費) 4,000,000 円、(間接経費) 1,200,000 円

研究成果の概要(和文)：2007年より始まった Linked Data ではメタデータそのものに焦点を当てている。メタデータの相互利用性を高めるためには、その記述規則を定義したメタデータスキーマと、応用ごとに決まるメタデータ語彙の組み合わせを定義したアプリケーションプロファイルに関する情報が有用である。しかしながら、現在の Linked Data においてメタデータスキーマに関する情報が十分に活用されているとは言えない。本研究では、Linked Data の柔軟な運用を目的としたメタデータスキーマに関する情報の流通と発見支援環境の作成をおこなった。

研究成果の概要(英文)：Many metadata schemas are used in the current Internet and Web environment. A conventional schema is designed in accordance with the requirements and standards for interoperability within a domain. However, cross-domain interoperability is still a problem. Even though there are some metadata schemas which are used commonly across domains, e.g. Dublin Core and FOAF, the barrier between domain-dependent schemas is a crucial problem from the viewpoint of metadata interoperability over the Internet. Dublin Core Application Profiles are intended to promote metadata interoperability across domains. It is well understood that metadata schema registries play an important role in promoting interoperability. This project

研究分野：メタデータ・デジタルライブラリ

科研費の分科・細目：情報学・図書館情報学・人文社会情報学

キーワード：メタデータ メタデータスキーマ Metadata Metadata Schema セマンティックWeb Semantic Web

1. 研究開始当初の背景

メタデータの記述方法の定義をメタデータスキーマと呼ぶ。メタデータスキーマではメタデータの記述に用いる属性語彙(例えば、タイトル、作者、更新日など)や属性値語彙(例えば、件名標目や分類番号など)の定義をおこなう。2001年頃から始まったWorld Wide Webコンソーシアムによるセマンティック Webの活動によってメタデータの利用はさらに広がってきた。これまでのセマンティック Webでは、主にメタデータスキーマ定義やメタデータの記述に用いるための語彙(例えば、RDF Schema, OWL, SKOSなど)の標準化がおこなわれてきた。さらに2007年春頃からメタデータの利用に焦点を当てたLinked Open Dataが開始された。LODではネットワーク上のRDFで書かれたメタデータを結びつけ、メタデータの相互利用性や流通性を高めることを目的としている。この結びつけられたデータセットを「Linked Data(リンクするデータ)」と呼ぶ。しかしながら、Linked Dataはメタデータに記述されたリンクを利用して情報資源間のつながりを密にし、メタデータの相互利用性を向上させているものの、メタデータスキーマで定義されている情報を有効利用できていない。今後のネットワーク上の情報資源の増加とセマンティック WebやLinked Dataにおけるメタデータへの需要の増加から、メタデータスキーマを効果的に利用することが、さらにメタデータの相互利用性を高めることになる。

2. 研究の目的

Linked Dataでは情報資源のURIを利用して、関連するメタデータの利用や発見をおこなう。URIはウェブページなどの情報資源に対してだけでなく、タグやラベルとして用いられる件名標目にも付与される。Linked Dataを利用したソフトウェアは、これらURIを利用して様々なサービスを実現しているが、利用するメタデータの構造(すなわちメタデータスキーマ)や記述されている値が用いている語彙が予めわかっていることを前提としているため、未知のメタデータスキーマで記述されたメタデータに対応することは難しい。そこで本研究では、Linked Dataの柔軟な運用を目的とするメタデータスキーマの流通・発見支援環境の作成をおこなった。本研究では、以下の3つの目標を設けて研究を進めた。

3. 研究の方法

(1)メタデータスキーマレジストリを基礎としたメタデータ記述項目発見支援モデルの提案と実現

様々なコミュニティが、それぞれの目的に合ったメタデータスキーマを作成し、メタデータを記述している。作成されたメタデー

タの長期利用や、コミュニティを超えたメタデータの利用を実現するためには、既存メタデータ語彙の共有と再利用を行い、メタデータの相互運用性を高めることが求められる。しかし、これらのメタデータスキーマはコミュニティ外での共有や相互運用を前提としていない場合が多い。また、新しいメタデータスキーマを作成する際に、既存メタデータ語彙からコミュニティの目的に合致し、再利用できる語彙を発見する手法は明らかではない。そこで本研究では、メタデータスキーマを作成する際のターム発見・選択における問題点を挙げ、それを改善し、メタデータスキーマ作成手順を示すメタデータスキーマ作成モデルを提案した。また、利用目的に合ったターム検索手法を提示し、メタデータスキーマ作成支援としてターム検索システムの開発を行った。

(2) Linked Dataのためのアプリケーションプロファイル生成モデルの提案と実現

記述されたメタデータ(メタデータインスタンス)は、既存のメタデータスキーマにコミュニティの目的に合わせたメタデータ語彙と制約(例えば、タイトルは日英で、件名は3件以上記述するなど)を組み合わせたことが多い。このような応用ごとに定義されるメタデータの構造をアプリケーションプロファイルと呼ぶ。アプリケーションプロファイルは、メタデータの構造と制約を理解するために重要であるにもかかわらず、明示的に定義されることは少ない。そこで本研究では、メタデータインスタンスを解析し、DCMI Description Set Profile形式に基づいたアプリケーションプロファイル生成モデルの提案と実現をおこなった。

(3)アジャイル開発に基づくメタデータスキーマ設計手法の提案と支援システムの実現

メタデータ作成者の要求要件を満たしたメタデータスキーマを設計するためには、何度もメタデータの試作とメタデータスキーマの修正を繰り返す必要がある。加えて、相互利用性の高いメタデータスキーマを設計するためにはメタデータ設計者に専門的な知識や経験が求められる。そのため、メタデータスキーマの設計作業には非常に手間がかかるという問題がある。また、第三者にとってメタデータを使いやすい状態にするためには、メタデータと共にメタデータスキーマも公開される必要がある。そこで本研究では、本研究において得たメタデータスキーマの設計に関する知見に基づき、メタデータスキーマの設計手法の提案とその手法に対する支援システムの開発をおこなった。これにより、メタデータスキーマ設計者がより効率的に要求に沿ったメタデータスキーマを設計することを目指した。

4. 研究成果

(1) 利用目的に応じたメタデータスキーマの作成手法の提案

長期に渡るメタデータの利用や、コミュニティを超えたメタデータの利用を実現するためには、既存メタデータ語彙の共有と再利用を行い、メタデータの相互運用性を高めることが求められる。しかし、これらのメタデータスキーマはコミュニティ外での共有や相互運用を前提としていない場合が多い。また、既存メタデータスキーマから、コミュニティの目的に合致し再利用できる語彙を発見する手法は明らかでない。そこで本研究では、メタデータ語彙の推薦手法を示し、タームの検索、語彙セット評価の過程を重要視したメタデータスキーマ作成モデルの提案を行った。本モデルは(1)要求分析、(2)構造構築・ターム検索、(3)語彙セット評価、(4)メタデータスキーマの記述、の4つのフェーズで構成される(図1参照)。

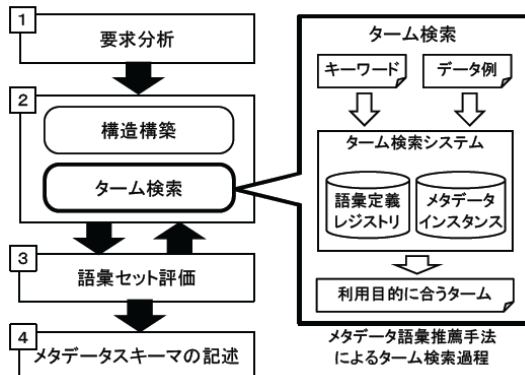


図1 利用目的に応じたメタデータ語彙推薦手法

(2) アプリケーションプロファイル作成支援を目的としたメタデータスキーマレジストリと LOD リソースに基づくメタデータ語彙の発見

次に本研究では、前述の提案モデルに従って、LOD リソースとして公開されているデータセットとメタデータ語彙を対象にしたメタデータ語彙発見の実験をおこなった。実験では、450 のメタデータ語彙と、LOD データセットとして公開されているメタデータインスタンス (RDF 記述で 3 億トリプル) を用いた。実験の結果、目的に応じたメタデータ語彙の発見には、既存メタデータ語彙とメタデータ語彙とメタデータインスタンスの情報だけでは十分ではなく、データセットのスキーマであるアプリケーションプロファイルに関する情報が必要であることがわかった。

(3) データセットのスキーマ分析とリンク関係の調査

メタデータスキーマ設計支援の精度を向上するため、既存のメタデータインスタンス

を調査し、メタデータの記述に用いられている語彙、スキーマ定義、データセット間の参照関係の調査をおこなった。調査は、CKAN 日本語 (151 データセット)、Open Data METI (201)、LOD チャレンジ Japan (120) を対象とした。

(4) 既存メタデータスキーマを用いたメタデータインスタンスからのスキーマ推定手法

本研究では LOD 利用者がデータセットのデータ構造を理解することを目的として、既存メタデータスキーマを用いたメタデータインスタンスからのスキーマ推定手法の提案と、本手法に基づいたシステムを構築した。本手法は、メタデータインスタンスから使用されているクラスとインスタンスを抽出し、公開されているメタデータ語彙定義からクラスとインスタンスの語彙定義を取得した上でスキーマの構成要素を推定することでスキーマを推定する。このとき、推定するために不足している情報は既に公開されているメタデータ語彙とアプリケーションプロファイルを利用して補完している。

(5) アジャイル開発モデルに基づくメタデータスキーマ設計手法と支援システム

メタデータ記述規則の設計プロセスはメタデータの試作とメタデータ記述規則の設計を繰り返すアジャイル開発モデルに基づくプロセスと捉えられる。しかしこのプロセスには、スキーマに従ったメタデータを試作するツールの開発に手間がかかる問題や、RDF のグラフ構造の構築やメタデータ語彙の探索が困難という問題がある。一方、メタデータを蓄積するためのデータベースは、そのスキーマにカラム名やデータ型など、グラフの構築やメタデータ語彙の探索の手がかりとなる情報を持つ。そこで本研究では、アジャイル開発モデルに基づく、メタデータ作成ツールの開発を通じたメタデータ記述規則の設計を行う手法とそれを実現する支援システムの提案をおこなった。本手法は、メタデータ作成ツールの開発を行いながらメタデータ記述規則の設計を繰り返し行う。本システムは、メタデータ作成ツールのデータベーススキーマの情報から RDF のグラフ構造の構築とメタデータ語彙の探索を行いメタデータ記述規則のひな形を生成した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計 8 件)

[1] 西出頼継, 本間維, 永森光晴, 杉本重雄. 日本の Open Data 活用を目的としたデータセットのスキーマ分析とリンク関係の調査, 第 112 回情報処理学会情報基礎とアクセス技術研究会, 2013 年 9 月 26 日, 筑波大学文京校

舎, 査読無

[2] Tsunagu Honma, Mitsuharu Nagamori, Shigeo Sugimoto. Find and Combine Vocabularies to Design Metadata Application Profiles using Schema Registries and LOD Resources, International Conference on Dublin Core and Metadata Applications 2013, 2013年9月2日-6日,TECNICO LISBOA ポルトガル, 査読有

[3] 落合香織, 三原鉄也, 永森光晴, 杉本重雄. DCMI Description Set Profileに基づくRDF Refineを利用したメタデータ作成支援手法の提案, 第29回セマンティックWebとオントロジー研究会, 2013年3月26日-27日, 伊豆大島大島温泉ホテル, 査読無

[4] 田中圭, 西出頼継, 永森光晴, 杉本重雄. Linked Open Dataを利用するアプリケーションの作成支援 - スキーマが与えられていないメタデータの視覚化支援手法の提案 -, 第75回情報処理学会全国大会, 2013年3月6日-8日, 東北大学, 査読無

[5] 小早川遥, 本間維, 永森光晴, 杉本重雄. 利用目的に応じたメタデータスキーマの作成支援 -メタデータ語彙の推薦と語彙セット評価手法の提案-, 第75回情報処理学会全国大会, 2013年3月6日-8日, 東北大学, 査読無

[6] Tsunagu Honma, Mitsuharu Nagamori, Shigeo Sugimoto. A Model to Support Interpretation of Embedded Metadata without Formal Schema by Linking a Metadata Instance to DCMI Description Set Profiles, International Conference on Dublin Core and Metadata Applications 2012, 2012年9月3日-7日, Pullman Hotel Kuching, マレーシア, 査読有

[7] 田中圭, 本間維, 永森光晴, 杉本重雄. Linked Open Dataを利用するソフトウェアの開発支援環境 メタデータスキーマレジストリを利用した記述規則の多様化への対応, 第74回情報処理学会全国大会, 2012年3月6日-8日, 名古屋工業大学, 査読無

[8] Mitsuharu Nagamori, Masahide Kanzaki, Naohisa Torigoshi, Shigeo Sugimoto. Meta-Bridge: A Development of Metadata Information Infrastructure in Japan, Proceedings of International Conference on Dublin Core and Metadata Applications 2011, 2011年9月21日-23日, National Library of the Netherlands, オランダ, 査読有

〔図書〕(計1件)

(1) 永森光晴 他、樹村房、図書館情報技術論、2014、6章「ネットワーク情報資源とメタデータ」、7章「図書館サービスにおけるネットワーク情報資源の利用」、p.208 (p.142-181)

6. 研究組織

(1) 研究代表者

永森 光晴 (NAGAMORI, Mitsuharu)

筑波大学・図書館情報メディア系・講師

研究者番号: 60272209