

令和 3 年 6 月 16 日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2018～2020

課題番号：18K11984

研究課題名(和文)メタデータスキーマを基礎とした持続的なメタデータ生成支援と流通支援環境の構築

研究課題名(英文)Construction of a sustainable metadata generation support and distribution support environment based on metadata schemas

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) 3,000,000円

研究成果の概要(和文)：政府、図書館、博物館などが作成したLODデータセットが、様々な形で利活用されている。しかし、あまり利活用されていないもの、更新が行われず古くなってしまったもの、また、公開が停止されアクセスできなくなってしまったものも少なくない。そこで本研究では、LODデータセットを長期に渡って公開し利活用性を高めるためには、メタデータのライフサイクルを支援する必要があると考え、次の目標を設けて研究を行った。

- (1)メタデータスキーマを基礎とした既存リソースからの持続的なメタデータ作成支援モデルの提案
- (2)LODデータセットの利活用性向上を目的としたメタデータスキーマを基礎とするメタデータ公開支援モデルの提案

研究成果の学術的意義や社会的意義

メタデータの重要性が認識され、多種多様な領域で膨大なメタデータが作成されるようになってきた。しかし、メタデータそのものを管理維持することができなければ、情報資源の発見や流通を支えるだけでなく、情報資源を長期に渡って理解して利用するという本来の役割を担うことはできない。本課題では、追加や修正を考慮した、メタデータの作成から管理運用までの過程を支援する仕組みと、メタデータの公開に際して利活用性を高めるための手法を明らかにした。

研究成果の概要(英文)：With the spread of Semantic Web technology and Linked Open Data (LOD), metadata created by various communities have been released as LOD datasets. LOD datasets created by governments, libraries, museums, etc. are used in various services and applications. However, there are many datasets that are not used much, are outdated, or are no longer accessible. In order to make metadata available to the public for a long period of time, and to improve its usability, this research considers the need to support the metadata lifecycle (schema design, metadata creation, metadata utilization, maintenance, and operation), and has set the following two goals

- (1) To propose a model to support sustainable metadata creation from existing resources based on metadata schemas.
- (2) To propose a metadata publication support model based on metadata schemas to improve the usability of LOD data sets.

研究分野：情報科学

キーワード：メタデータ メタデータスキーマ セマンティックWeb Linked Open Data

1. 研究開始当初の背景

メタデータは「データに関するデータ」と定義され、情報資源の組織化や検索のためだけではなく、保存や流通のための重要な役割を担うようになって来た。現在では、学会、博物館、政府、地方自治体といった様々なコミュニティが目的に合わせたメタデータを作成している。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）と属性値語彙（例えば、件名標目や分類番号など）の定義を行う。2001年頃から始まったセマンティック Web の活動によってメタデータの利用はさらに広がっている。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(Web Ontology Language), SKOS(Simple Knowledge Organization System)など）の標準化が行われてきた。さらに2007年頃からメタデータの利用に焦点を当てた W3C による LOD プロジェクトが開始された。LOD プロジェクト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)」と呼び、その中でもオープンなライセンスで公開されているものを LOD (Linked Open Data) データセットと呼ぶ。2013年の G8 ロックアーンサミットにおけるオープンデータ憲章の合意以降、公開される LOD データセットの数は増加の一途をたどっており、利活用性向上が課題となっている。

メタデータの重要性が認識され、多種多様な領域で膨大なメタデータが作成されるようになってきた。しかし、メタデータそのものを管理維持することができなければ、情報資源の発見や流通を支えるだけでなく、情報資源を長期に渡って理解して利用するという本来の役割を担うことはできない。本課題では、追加や修正を考慮した、メタデータの作成から管理運用までの過程を支援する仕組みと、メタデータの公開に際して利活用性を高めるための手法を明らかにする。

2. 研究の目的

様々なコンテンツが企画設計、制作、利用、保存管理という段階を持つように、メタデータもライフサイクルを持つ。メタデータの場合、まず作成者はメタデータの記述対象である情報資源の特徴を分析し、利用目的に合わせて適切なメタデータ語彙を選択してメタデータスキーマの設計を行う。そして、そのメタデータスキーマに従ってメタデータを作成する。メタデータを一から作成するだけでなく、既存のリソース（例えばデータベース、Excel、ウェブページ等）を変換してメタデータ作成に利用することも多い。公開されたメタデータは利用者によってサービスやアプリケーションの中で利活用される。必要に応じてメタデータスキーマ及びメタデータは追加・修正されて維持・運用される。

セマンティック Web 技術と LOD の普及に伴って、多種多様なコミュニティが作成したメタデータが LOD データセットとして公開されるようになってきたが、メタデータ作成と維持には手間とコストがかかる。そのため、LOD プロジェクトの活動が始まってから10年が経ち、LOD データセットの中には更新が行われずメタデータが古くなってしまったものや、あまり利活用されていないもの、また公開が停止されデータセットそのものにアクセスができなくなってしまった事例も少なくない。また多くの場合、メタデータ提供者は、そのコミュニティの扱う領域の専門家であって、メタデータ語彙やメタデータスキーマについての専門知識やメタデータスキーマ設計に関する経験、そして LOD データセットを公開し運用するための知見を持っているとは限らない。そこで本研究では、メタデータを長期に渡って公開し、利活用性を高めるためには、メタデータのライフサイクル（スキーマ設計、メタデータ作成、メタデータ利活用、維持・運用）を支援する必要があると考え、以下の2つの目標を設けて研究を進めた。

- (1) **メタデータスキーマを基礎とした既存リソースからの持続的なメタデータ作成支援モデルの提案**：メタデータは、既存リソース（データベース、Excel、CSV、ウェブページ等）の変換や抽出によって生成されることが多い。これらリソースは、日々の業務の中で追加や修正が行われ変化してゆく。そこで本研究では、リソースの追加・修正に合わせてメタデータを更新するためのメタデータ作成支援モデルの提案を行なった。本モデルでは、メタデータスキーマと多様なリソースのマッピング手法と、メタデータスキーマに基づくリソースの変化を考慮した持続的なメタデータの作成手法の提案を行なった。
- (2) **LOD データセットの利活用性向上を目的としたメタデータスキーマを基礎とするメタデータ公開支援モデルの提案**：公開されている LOD データセットの中には、利活用性が低いものも少なくない。この原因のひとつは、データセットの公開（アーカイブされたファイル、問い合わせのための SPARQL エンドポイント）に際して、そのメタデータスキーマ（記述に利用しているメタデータ語彙やメタデータの構造）が十分に説明されていないためである。本研究では、LOD データセットの利活用性向上を目的として、メタデータスキーマを利用した LOD データセットで利用されているメタデータ語彙と構造情報の理解支援モデルを提案を行った。

3. 研究の方法

まず平成 30 年度において、メタデータスキーマとリソース中の関連する項目を結びつけるための抽出ルール記述形式の検討を行った。そして、LOD データセットの利活用性を向上させるために、メタデータに関する知識が少ない利用者がメタデータ記述に使われている語彙やスキーマを理解するための要求要件とその提示方法について検討を行った。その後、既存リソースからのメタデータ抽出と理解支援のためのプロトタイプを作成し、その評価を行った。

令和元年度では、平成 30 年度での評価に基づき、メタデータスキーマと既存リソースのマッピング手法とシステムの改善を図った。そして、メタデータの生成から公開までの一連の過程を支援するためのモデルの提案と、そのためのシステム構築とその評価を行った。

令和 2 年度は、前年度までの評価に基づいて、マッピング手法とシステムの改善を行った。

4. 研究成果

(1)メタデータスキーマを基礎とした既存 WebAPI からの持続的なメタデータ作成支援

Web 上の情報が増加してきた現在、計算機が理解可能な形で情報を公開できる Linked Open Data(LOD)が注目されている。LOD として公開されているデータセットは SPARQL を用いて検索することができ、さらに横断検索機能である Federated Query を用いることで複数の LOD データセットに問い合わせ、それぞれの結果を組み合わせた情報を得ることができる。LOD は様々な情報と組み合わせることでさらに利活用を広げている。

一方で、Web 上では LOD データセットとして存在しない情報が多く提供されている。例えば、検索サービスの Web API を活用することで天気などの動的データを取得できる。本研究では RDF 以外の形式で公開されているデータを非 LOD と呼ぶ。非 LOD と LOD の情報を組み合わせたアプリケーションを開発する時、それぞれの情報を別々に取得し、個別に組み合わせる必要がありコストがかかってしまう。SPARQL 式を通して Web API などが提供している非 LOD に問い合わせることができれば、LOD としては扱いづらい動的データとの連携も容易になり、LOD の利活用性の向上が期待できる。本研究では、Web API から取得した非 LOD と既存の LOD を繋げることを目的として、SPARQL 式から Web API へ問い合わせることができる環境の構築を行った。

本研究では Web API を対象に、スキーママッピングを用いて非 LOD の取得と RDF グラフの生成を自動的に行うシステムを構築した。WebAPI 利用者はマッピングを記述することなく、SPARQL 式を用いて Web API への問い合わせを実行でき、事前に用意されたマッピングに基づいて LOD 化された結果が取得できる。多くの Web API が問い合わせ方法や取得できるデータ構造などのメタデータスキーマを公開している。このメタデータスキーマを用いて Web API への問い合わせの設定および、取得データと RDF データモデルのマッピングを事前に用意することができ、自動的に行うことができる。このマッピングをスキーママッピングと呼ぶ。Web API から非 LOD を取得し LOD 化するまでの手順を以下に示す。

Web API への入力値の取得：Web API にアクセスし、非 LOD を取得するために、利用者の記述した SPARQL 式から問い合わせに必要な入力値を取得する。

Web API への問い合わせ実行：Web API ごとに決められた方法で URL を生成し検索を実行する。

RDF グラフの生成：取得した非 LOD から値を取り出し、それぞれの値を目的語として RDF グラフに格納する。

Web API の公開している情報からこれらの情報を取り出しスキーママッピングファイルとして用意しておくことで、それぞれの手順で必要な情報をそこから読み取り自動的に実行していくことができる。

多くの情報が LOD データセットとして公開されている一方で、LOD データセットとして公開されていない情報が Web 上に多く存在する。しかし、これらの非 LOD と LOD を組み合わせて利用するためにはコストが大きく、効率的に双方の情報を取得して結合できる環境は整っていない。これに対して本研究では Web API のメタデータスキーマを用いて RDF データモデルとのマッピングを事前に記述することで非 LOD と LOD を統合する手法を提案した。また、提案手法に基づいたシステムを構築し、SPARQL 式から Web API へ問い合わせによる持続的なメタデータの生成を行った。

(2) LOD データセットの利活用性向上を目的としたメタデータスキーマを基礎とするメタデータ公開支援

LOD データセットは RDF 形式でメタデータが記述され、外部へのリンクを値として含むことができるため、他のデータセットと組み合わせ利用しやすく、再利用性も高い。また、様々な組織の取り組みによって、公開される LOD データセットの数は年々増加している。そのため、既存の LOD データセットの様々な利活用が期待されている。しかし、LOD データセットの公開数の増加に反してその利用は活発に行われていない。この要因の一つとして、データセットの探索・利用時に用いられるデータセットに関するメタデータ群が十分に整備されていないことが挙げられる。本研究では LOD データセット利活用性の向上を目的として、データセットの探索から利用

までを支援するデータセットに関するメタデータ群を記述する、メタデータプロファイルを定義し、その作成・提供を支援するシステムを構築を行った。更に、提案システムによって半自動的に登録されたメタデータプロファイルのデータセット利用時の構造理解における有用性を示すための評価実験を行った。

LOD データセットの探索・利用には、そのデータセットの仕様に関するメタデータが求められる。例えば、データセットの作成者や更新状況に関する情報はデータセットの信頼性や品質を表し、利用するデータセットを選択する上での指標となる。また、データセットのクラス・プロパティ定義や、プロパティの記述制約といった構造に関する情報はデータの利用を容易にする。しかし、現在提供されるデータセットに関する情報は十分に整備されていない。利用に必要な情報が公開されていない場合は、利用者が独自にデータセットの中身を参照し調査を行う必要があるが、LOD データセットは人間にとっての可読性が低いため、調査の手間とメタデータに関する専門知識を要する作業となる。そのため、データセットに関する情報の整備不足が LOD データセット利活用の妨げとなっている。

そこで本研究では、データセットの探索から利用までの内容理解を支援する情報群としてメタデータプロファイルを定義した。メタデータプロファイルは、データセットの内容理解と利用の支援、 目的に適したデータセットの発見、それぞれの支援を目的とする。

現在公開されるデータセットに関する情報は、人間・コンピュータが双方に理解できるメタデータスキーマの公開が進んでおらず、また利用者が持ちうる利用方法や利用例に関する情報は共有されていない。そのため内容理解、用法理解を充分支援できるほど情報が整備されていない。そこで本研究で定義するメタデータプロファイルでは、人間可読な情報を含む DCMI シンガポールフレームワークに基づくメタデータスキーマ、データセットに関する情報を中心に定義する。また、データセットを利用者が発見しやすくするための項目を与え、キーワードを用いた検索を行えるようにする。以上より、メタデータプロファイルは主に以下のような項目から構成する。

(a) データセットの基本情報に関する項目

データセットのタイトル、概要、作成者など、データセットの発見・分類に有用な書誌的基本項目を与える。項目の値に含まれるキーワードからデータセットの検索を行えるようになり、データセットの発見支援に用いる。

(b) メタデータスキーマに関する項目

LOD データセットにおいて、どのような情報がどのように記述されているのかといった、データセットの構造に対する詳細な情報が与えられることで、利用時のデータアクセスを円滑に行える。本研究では利用者がより構造を理解しやすくするための以下の項目を整備した。

・ドメインモデル

クラスをノード、クラスリソース間をリンクさせるプロパティをエッジとして、記述対象のリソース間の関係を描いたグラフをドメインモデルと呼ぶ。ドメインモデルを与えることで、主要なリソースの分類やその間の関係性を一目で把握することができ、構造の骨子の理解に役立つ。

・記述規則の項目

記述対象リソースのクラスと、クラス毎のリソースが持つプロパティに関する項目を与える。URI やラベル、プロパティ値のデータ型などの制約を与える。また、内容理解を支援する人間可読な構造情報として、プロパティ毎にデータセット中で実際にそのプロパティを用いているトリプルの例(以下、トリプル例と呼ぶ)を与える。各プロパティの具体的な用法の定義はコストがかかるが、簡易なトリプルで表現される用法例を与えることで、利用者がプロパティの意味をある程度理解できるようにする。

・記述規則に関する利用者提供の項目

構造情報は、データセットを構成する一つ一つのクラスとプロパティに対し付与されることで利活用時に有用となる。しかし、データセット公開者では十分な情報の整備ができない場合がある。また、有用な利用情報を持ちうるデータセット利用者の情報は共有されていない。そこでデータセットの利用者が独自に持つ情報を記述規則の細かい粒度の項目として与えられるよう定義し、利用者情報の共有の枠組みを作る。これにより、構造情報をより充実させ、内容理解・用法理解の支援を行う。

(c) データセットの利用例に関する項目

利用例はデータセットの機能要件の理解の指針となる。そこで、データセットを利用したアプリケーションや利用に関連するウェブ上の記事についての項目を与える。データセットの公開者だけでなく、データセットの利用者からも情報を収集できることで、より多様なデータセット利用の方針を提示でき、LOD データセットの発見支援となる。

(d) データセットに対する自由記述項目

データセット利用者は様々な観点からの情報を持ちうると思われ。利用者からのより多様な情報を収集するために上記の項目には含まれないデータセットに関する情報を wiki のよう

に自由に記述できる項目を設ける。これによって情報提供者の持つ多様な情報の収集に対応する。

本研究では LOD データセットの利活用性の向上を目的として、LOD データセットの探索から利用までを支援するメタデータプロファイルを定義し、その公開支援を行うシステムを構築した。メタデータプロファイルはデータセットの内容理解、用法理解を促進するために人間向けの構造情報や利用者視点の情報を要素として含み、本システムによって構造情報の半自動的な分析や、ブラウザ上で利用者・公開者双方による容易な登録・編集を行なうことができる。

実験では、本研究で定義したメタデータプロファイルの半自動的に登録できる構造情報の有用性を評価した。結果として、ドメインモデルとトリプル例を表形式の記述規則とともに示すことで、SPARQL クエリ作成時の構造理解・語彙の用法理解に効果があった。しかし、利用者がデータセットの扱う領域の知識がない場合は、ドメインモデルとトリプル例では十分な理解が得られず、人間向けの語彙の詳細な定義、用法に関するコメントなどが必要である。またシステムでのメタデータプロファイルの表示もより直感的にわかりやすく情報を閲覧できるよう改良が求められることがわかった。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計15件（うち招待講演 0件 / うち国際学会 7件）

1. 発表者名 Thalhath, N., Nagamori, M., Sakaguchi, T., Kasaragod, D., Sugimoto, S
2. 発表標題 Semantic Web Oriented Approaches for Smaller Communities in Publishing Findable Datasets
3. 学会等名 Research Conference on Metadata and Semantics Research (MTR 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Thalhath, N., Nagamori, M., Sakaguchi, T., & Sugimoto, S.
2. 発表標題 Wikidata Centric Vocabularies and URIs for Linking Data in Semantic Web Driven Digital Curation
3. 学会等名 Research Conference on Metadata and Semantics Research (MTR 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Thalhath, N., Nagamori, M., & Sakaguchi, T.
2. 発表標題 MetaProfiles A Mechanism to Express Metadata Schema, Privacy, Rights and Provenance for Data Interoperability
3. 学会等名 International Conference on Asian Digital Libraries (ICADL 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 岩波春香, 永森光晴, 三原鉄也
2. 発表標題 メタデータプロフィールを利用したLODデータセット利活用支援
3. 学会等名 第53回セマンティックウェブとオントロジー研究会
4. 発表年 2021年

1. 発表者名 大里摘実, 三原鉄也, 永森光晴
2. 発表標題 既存のWebリソースを用いたアニメ作品のスタッフ情報LODの構築
3. 学会等名 第5回コミック工学研究会
4. 発表年 2021年

1. 発表者名 半澤輝尚, 三原鉄也, 永森光晴
2. 発表標題 構造記述を利用したマンガのメタデータ分析基盤 -マンガの研究データのLOD化 -
3. 学会等名 第5回コミック工学研究会
4. 発表年 2021年

1. 発表者名 糸井峻, 三原鉄也, 永森光晴
2. 発表標題 マイクロタスクを用いたコマ画像の再構成によるマンガの要約
3. 学会等名 第5回コミック工学研究会
4. 発表年 2021年

1. 発表者名 Kosuke Oishi, Tetsuya Mihara, Mitsuharu Nagamori, Shigeo Sugimoto
2. 発表標題 Identifying and Linking Entities of Multimedia Franchise on Manga, Mani and Video Game from Wikipedia
3. 学会等名 Digital Libraries at Crossroads of Digital Information for the Future (Springer LNCS 11853) (国際学会)
4. 発表年 2019年

1. 発表者名 Nishad Thalhath, Mitsuharu Nagamori, Tetsuo Sakaguchi, Shigeo Sugimoto
2. 発表標題 Authoring Formats and Their Extensibility for Application Profiles
3. 学会等名 Digital Libraries at the Crossroads of Digital Information for the Future (Springer LNCS 11853) (国際学会)
4. 発表年 2019年

1. 発表者名 Nishad Thalhath, Mitsuharu Nagamori, Tetsuo Sakaguchi, Shigeo Sugimoto
2. 発表標題 Yet Another Metadata Application Profile (YAMA): Authoring, Versioning and Publishing of Application Profiles
3. 学会等名 Proceedings of International Conference on Dublin Core and Metadata Application 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Nishad Thalhath, Mitsuharu Nagamori, Tetsuo Sakaguchi, Shigeo Sugimoto
2. 発表標題 Metadata Application Profile Provenance with Extensible Authoring Format and PAV Ontology
3. 学会等名 Joint International Semantic Technology Conference 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 武井 響也, 三原 鉄也, 永森 光晴, 杉本重雄
2. 発表標題 マイクロタスクによるマンガの暗黙構造についてのメタデータ作成 - マンガのコマを読む順序とテキストの話者について -
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2019年

1. 発表者名 山中勇樹, 三原 鉄也, 永森 光晴, 杉本重雄
2. 発表標題 関連するLODデータセットの発見を目的としたリンク関係の調査
3. 学会等名 第47回セマンティックウェブとオントロジー研究会
4. 発表年 2019年

1. 発表者名 阪本かすみ, 三原 鉄也, 永森 光晴, 杉本重雄
2. 発表標題 スキーママッピングを基礎としたWeb APIとLODの統合的利用環境の構築
3. 学会等名 第47回セマンティックウェブとオントロジー研究会
4. 発表年 2019年

1. 発表者名 志賀大輝, 三原 鉄也, 永森 光晴, 杉本重雄
2. 発表標題 書誌情報と主題語彙のリンキングによるマンガの主題検索のためのデータセット構築
3. 学会等名 第47回セマンティックウェブとオントロジー研究会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------