

# 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—

小早川 遥<sup>†</sup> 本間 維<sup>‡</sup> 永森 光晴<sup>†† ‡‡</sup> 杉本 重雄<sup>†† ‡‡</sup>

筑波大学情報学群情報メディア創成学類<sup>†</sup> 筑波大学大学院図書館情報メディア研究科<sup>‡</sup>

筑波大学図書館情報メディア系<sup>††</sup> 知的コミュニティ基盤研究センター<sup>‡‡</sup>

### 1. はじめに

W3C によるセマンティック・ウェブでは、リソースに機械可読なメタデータを付与することで、Web of Data を実現するための試みが行われてきた。そうした試みの中で、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。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を提案する。

### 2. メタデータスキーマ

図1は書籍のメタデータ記述例である。この例では、書籍のタイトルや著者名といったメタデータ記述項目に dc:title や foaf:name などのタームを用いている。これらのタームは、メタデータ語彙である DCMI Metadata Terms<sup>[1]</sup>や FOAF 語彙<sup>[2]</sup>の中で定義されている。メタデータスキーマは、メタデータ記述で使用するタームの集まりをメタデータ語彙として定義し、応用ごとに決まるメタデータ記述の構造や各記述項目の制約を記述規則として決めている。図1のように幾つかの既存メタデータ語彙を再利用し、タームを組み合わせてメタデータを記述することで、共通語彙によるメタデータの横断利用が可能になり、メタデータの相互運用性を高めることができる<sup>[3]</sup>。DCMI<sup>[4]</sup>では、既存メタデータ語彙の再利用を考慮したメタデータスキーマモデルとして Dublin Core Application Profile を提唱している。

### 3. メタデータスキーマ作成における問題

現在、新しいメタデータスキーマを作成する際に、目的に合致した既存メタデータ語彙を発見する具体的な手法やメタデータスキーマ作成手順は確立されていない。Application Profile に基づくメタデータスキーマの作成では、(1)要求分析・定義、(2)ドメインモデルの構築、(3)メタデータ記述項目定義、(4)エンコーディング方式策定、という手順を提案している<sup>[5]</sup>。しかしながら、これは非常に大まかな手順であり、これに基づきメタデータスキーマを作成していく事は難しい。実際に幾つかのメタデータスキーマを作成し分析したところ、手順2と3の過程はほぼ同時に行われ、また多くの知識を必要とし、時間のかかる過程であることがわかった。手順2と3において特にコストが大きいのは、既存タームを探す作業と、使用するタームを決定する作業である。タームを探し決定する作業にコストがかかる原因として次の問題が考えられる。

- (a) 既存のメタデータ語彙を収集し利用できる環境が整っていない
- (b) 利用目的に合うタームを探す手段がない
- (c) 相互運用性や利用目的を考慮した語彙セットを選択するための評価基準がない

本研究では、上記の問題を解決するために、ドメインモデルの構築とターム検索を行うフェーズと、語彙セット評価のフェーズを用意したメタデータスキーマ作成モデルの提案をする。

### 4. メタデータスキーマ作成モデルの提案

図2に本研究で提案するメタデータスキーマ作成モデルの概要を示した。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。

第1フェーズでは、作成するメタデータの利用目的を明確にし、記述項目や構造を大まかに考える。第2フェーズでは、構造構築とターム検索を同時に行う。ターム検索には、5章で述べるメタデータ語彙推薦手法に基づき開発したターム検索システムを

“Development of a Support Environment for Making Metadata Schemas based on User Requirements”

<sup>†</sup>Haruka Kobayakawa. School of Informatics, Univ of Tsukuba.

<sup>‡</sup>Tsunagu Honma. Graduate School of Library, Information and Media Studies. Univ of Tsukuba.

<sup>††</sup>Mitsuharu Nagamori. Shigeo Sugimoto. Faculty of Library, Information and Media Science, Univ of Tsukuba.

<sup>‡‡</sup>Reserch Ctr for Knowledge Communities. Univ of Tsukuba.



図1 メタデータ記述例

利用する。第3フェーズとなる語彙セット評価では、記述対象の種類ごとにタームセットを評価する。評価が低い場合や次の記述対象分類のターム検索に移る場合に第2フェーズに戻り、構造を確定しながら使用するタームを決定する。第4フェーズで、語彙定義や記述規則を含めたメタデータスキーマを記述し、メタデータスキーマを完成させる。

前章で述べた(a)~(c)の問題は、それぞれ本モデルにおける語彙定義レジストリ、ターム検索システム、語彙セット評価の3要素によって解決する。(a)の問題を解決する語彙定義レジストリとは、本研究においてDCMI Metadata Termsなどの代表的メタデータ語彙定義約130件を収集し、蓄積・利用できるようにしたものである。ターム検索システムは、語彙定義レジストリを用い、5章に述べるターム検索手法をシステムとして実現したものである。これは(b)の問題を解決する。また、第3フェーズで語彙セットの評価基準を定め、タームの組み合わせ評価を明確にすることで、(c)の問題を解決する。

## 5. 利用目的に合ったメタデータ語彙推薦手法

メタデータスキーマ作成モデルのターム検索過程におけるメタデータ語彙推薦手法について述べる。

### 5.1. キーワードによるターム検索手法

キーワードによるターム検索には、メタデータ記述項目を表す言葉とその関連語をキーワードとして利用する。例えば、楽曲アルバムのメタデータとしてアルバムの「トラック」を表す項目を記述したい場合、「track」といった項目を表す言葉や、「album」、「song」などの関連語がキーワードとなる。これらを入力として与え、語彙定義レジストリに登録されている語彙定義からターム名やその説明文にキーワードを含むものを結果として返す。図3では「album track」を入力として与えた結果を示している。また、この検索手法では、曲名を表す「title」と称号であることを指す「title」は意味が異なるが区別なく検索されてしまう。そこで、語彙をGeneral、Library、City、Web、Where&When、

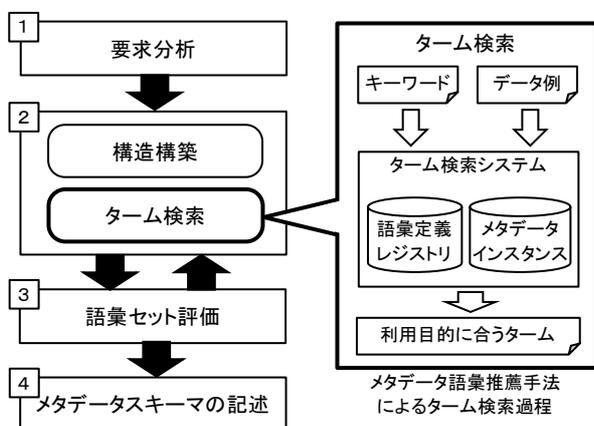


図2 4つのフェーズで構成されるメタデータスキーマ作成モデル

Science、Media、Marketの8つのカテゴリに分類することで、特定のカテゴリに限定した検索を行う。

検索結果は、語彙定義にキーワードが多く含まれているものを上位に推薦するのではなく、メタデータを共有するためのサービスであるLOD Cloud Cache<sup>[6]</sup>に集められたメタデータインスタンスを利用し、タームの使用率が高い順に推薦する。そうすることで相互運用性の高いタームの選択が可能となる。

### 5.2. データ例によるターム検索手法

データ例によるターム検索手法では、データ例を実際に値として持つタームを検索する。データ例とはメタデータ記述項目の値の記述例のことである。例えば、記述項目「曲名」の値のデータ例は「Obladi Oblada」などの実在する曲名になる。これを入力として与え、メタデータインスタンスの中から、与えたデータ例を値として持つタームを探索する(図3)。

キーワードによるターム検索では、例えば「singer」と「artist」のように同じ意味でも表現の違いにより語彙定義文とキーワードが一致しない場合がある。しかし、データ例によるターム検索ではデータ例を実際に値として持つタームを検索結果とするため、言葉の表現に差異が生じる場合やメタデータ記述項目を表すキーワードがはっきりと示せない場合でもタームを探すことができる。

## 6. おわりに

本論文では、語彙定義レジストリ、ターム検索システム、語彙セット評価を要素とし、メタデータスキーマ作成過程における問題を解決するメタデータスキーマ作成モデルを提案した。これにより、メタデータスキーマ作成の道筋が明確化し、利用目的に合ったタームの検索が可能になった。今後は、ターム検索の精度の向上と今回論ずることのできなかつた語彙セット評価手法を明らかにする。

### 参考文献

- [1]DCMI Metadata Terms. <http://dublincore.org/documents/dcmi-terms/>
- [2]FOAF Vocabulary Specification. <http://xmlns.com/foaf/spec/>
- [3]メタデータ情報基盤構築事業、メタデータ情報共有のためのガイドライン. <http://meta-proj.jp/A03.pdf>
- [4]Dublin Core® Metadata Initiative (DCMI). <http://dublincore.org/>
- [5]谷口祥一.メタデータの「現在」.勉誠出版,2010
- [6]LOD Cloud Cache. <http://lod.openlinksw.com/>

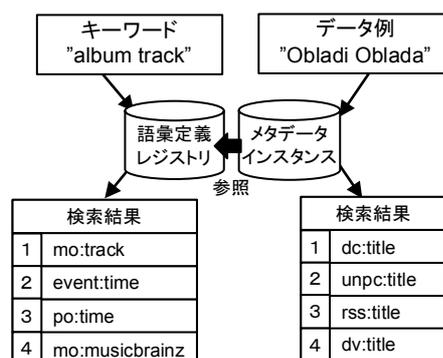


図3 ターム検索結果例