

Mahara における研究成果物のメタデータ管理手法 A metadata management method of artifacts in Mahara

西村 友基[†] 佐野 雅彦[†] 松浦 健二[†] 谷岡 広樹[†]
Yuki Nishimura Masahiko Sano Kenji Matsuura Hiroki Tanioka

大平 健司[†] 関 陽介[†] 上田 哲史[†]
Kenji Ohira Yosuke Seki Tetsushi Ueta

1. はじめに

大学における研究活動では、論文などの公開研究成果だけでなく、研究室に蓄積された非公開の研究成果も活用して、活動が推進される。その際、生成される実験結果や調査データ、論文や研究発表などの増加する研究成果を、後続する研究活動に効率的に活用するための有効な蓄積・共有方法が求められる。

e ポートフォリオシステムは個人の活動情報を集約し、振り返りや他者との成果物共有などを可能とするものである。この e ポートフォリオシステムを研究活動に適用することにより研究活動の効率化が期待できる。一般に、e ポートフォリオシステムでは、収集データにメタ情報を付与することで、メタ情報を用いた検索を可能とする。例えば Mahara[1]にあるデータ（以後アーティファクトと呼ぶ）には、作者、作成日、タイトル等の情報が自動付与され、内容説明として自由記述の情報を追加付与できる。検索は、アーティファクトを対象に行われ、指定した検索キーワードに該当するアーティファクトや追加付与した情報を単純に表示するだけである。そのため、関連性の高い収集データが上位に来るとは限らない。これらの検索結果をすべて確認するのは不可能ではないが、収集データに付与された情報と、他の収集データとの関係が把握できると、より柔軟で効率的な収集データの検索が期待できる。野村らは、成果物間の関係性と類似度を用いた検索結果を提示するシステムを提案している[2]。収集データのメタ情報同士の関係性を構築し、それを解析・提示することができれば、研究者間での知識・成果の共有の効率化が期待できる。

本研究では、Mahara に対し管理データベースシステム機能を拡張することにより、研究成果物へのメタ情報付与とそれらの管理の実現する。

2. 関連研究

公開論文のメタ情報を利用した成果物の管理に関する研究は複数行われている。廣川らは、論文のメタ情報を形態素解析したデータを用いた論文検索システムを提案している[4]。Cuong らは、論文のメタ情報をクラスタリングしたデータを用いた研究者の研究履歴作成を行っている[5]。横らは、論文のメタ情報から科学技術の時系列的な動向を分析している[6]。このように公開論文の場合は論文情報が公開されている場合が多いため、その情報を取得・解析・検索に用いればよい。一方、非公開の研究成果物の場合は、研究成果物に関する情報の付与が定型化していないため、解析・検索などに手間を要する。そこで、非公開の研究成

果物から抽出したメタ情報を付加し管理することで、解析・検索にかかる手間を減らす。

また、研究成果物の再利用を目指した成果物の蓄積に関する研究も行われている。森本らは、e ポートフォリオ間の関係を記述する文法を開発し、e ポートフォリオの再利用を支援している[7]。中村らは、学生の成果物を企業や教員が評価した情報を成果物のメタ情報として用いて、成果物を抽出する手法を提案している[8]。

3. 提案手法

3.1 Resource Description Framework

Resource Description Framework (RDF)[3]は、標準的なメタ情報記述のための枠組みである。Web ページのタイトルや更新日などの Web 上の資源に関する情報の記述に用いる。RDF は、メタ情報を「主語」「述語」「目的語」に相当するトリプルと呼ばれる 3 要素の組で表現する。例えば「A のタイトルは B である」という関係は「A title B」という 3 つ組で表現される。このとき、語彙はスキーマを記述する言語 (RDF Schema や OWL) で意味と構造を定義することで、異なるコミュニティ間で共有可能である。また、コンピュータが語彙の関係や意味に基づいた処理をすることが可能である。

RDF でデータ公開を行う組織は複数存在しており、例えば、CiNii はウェブ API により、論文情報を RDF データで提供している[9]。また、DBpedia は Wikipedia の情報を RDF データで提供している[10]。

3.2 提案手法の流れ

提案手法では、Mahara にあるアーティファクトを対象にメタ情報を付与する。具体的には、アーティファクトの作

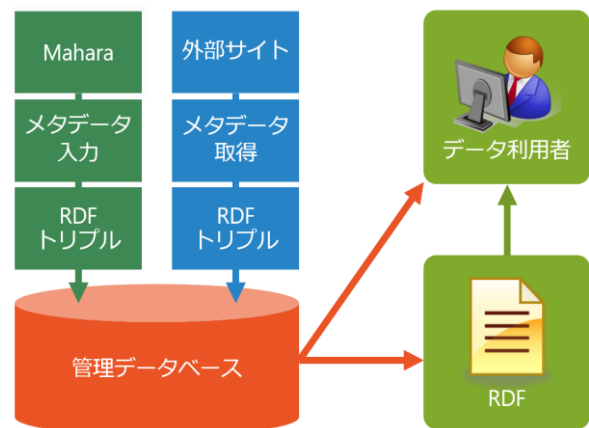


図 1 本手法の流れ

[†] 徳島大学
Tokushima University

者, 作成日, タイトル等の情報と, 研究成果物の種類に応じた研究活動に有用な情報を追加したメタ情報を Mahara の管理データベースとは異なる管理データベースに蓄積する. メタ情報は RDF に基づいて表現され, Web 上に公開された他のメタ情報とも連携する. 管理データベースに保存されたメタ情報は, データ利用者に管理データベースの直接参照または RDF/XML ファイル形式で提供する. 提案手法の流れを図 1 に示す.

3.3 研究成果物のメタ情報

メタ情報は, RDF の考えに基づき表現する. RDF では, メタ情報の記述に使用する語彙の定義を必要とするが, 本研究では, 表 1 表に示す語彙の名前空間を用いる. これには, Dublin Core (表 1 の dc と dcterms) や FOAF (表 1 の foaf) などの広く一般に使用されている語彙集を含めることにより, 他の RDF データとの互換性を考慮している.

表 1 名前空間

接頭辞	名前空間
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
dc	http://purl.org/dc/elements/1.1/
dcterms	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
org	http://www.w3.org/ns/org#
bibo	http://purl.org/ontology/bibo/
cinii	http://ci.nii.ac.jp/ns/1.0/
ndctype	http://ndl.go.jp/ndctype/

3.4 メタ情報管理

3.4.1 データベース設計

本研究では, メタ情報を格納するための管理データベースとして従来からある SQL データベース (実験システムでは MySQL) を用いた. この理由を以下に示す.

- Mahara 側で利用されているデータベースシステムを流用し, システム構成の複雑化を避ける.
- 特定データベースシステムへの依存度を抑えた設計が可能となる.

RDF ではトリプルですべての情報が表現されるため, 管理データベースでは 1 トリプルを 1 レコードとして取り扱い, レコード内の 3 つのカラムにトリプル内の情報をそれぞれ格納する方式を用いた. よって, 新しく属性が追加される場合でも, カラムを追加するといったテーブルスキーマの変更が不要となっている.

また, 情報の変更履歴をレコード単位で追跡可能とするため, メタ情報を更新する際は新しくレコードを登録するものとし, 更新後レコードから更新前レコードを参照するためのカラムを設けることとした. 表 2 にメタ情報を格納するデータベースのテーブル構造を示す.

3.4.2 メタ情報の付与

Mahara に蓄積保管する研究成果物をいくつかのカテゴリに分類して, それぞれにメタ情報の属性を定義する. 現段階では, 論文, 発表スライド, 議事録を対象とするが, 後述するテンプレート機能により比較的簡単にカテゴリを追

表 2 メタ情報格納テーブル

カラム名	内容
id	メタ情報の ID (主キー)
artifact_id	研究成果物の ID
s	RDF トリプルの主語
p	RDF トリプルの述語
o	RDF トリプルの目的語
id_old	更新前のメタ情報の ID
enabled	有効または無効

加可能としている. 表 3 は, メタ情報付与の対象が論文の場合の属性セットの例である. アクセス権は, 属性毎のアクセス権が設定された情報であり, そのアクセス制御には, Mahara 上のアーティファクト共有におけるアクセス制御を利用する. この方法により, 管理データベース上のアクセス制御を Mahara 側と一体化でき, 利用者から見たアクセス管理の複雑化を避ける効果が期待できる.

メタ情報は, Mahara にアーティファクトを投稿する際に, 自動・手動で付与する. 作者や作成日, タイトル等は, Mahara から自動で抽出され, 入力フォームに自動入力される. その他のキーワードや要約等は手動入力する. 図 2 は Mahara 上に実装したメタ情報入力画面である. 成果物の種類を選択するとその属性セットの入力フォームが生成される. 入力されたメタ情報は RDF に変換し管理データベースに保存する.

研究成果物はカテゴリ毎にそれぞれが必要とする属性情報をテンプレート化し定義している. そのため, 研究成果物の属性を追加・変更する場合は, このテンプレートに属性の追加とフォーム処理の追加により対応できる. カテゴリを追加する場合も同様である.

3.4.3 外部情報との連携

本研究では, オープンデータとして公開されている研究成果物に関する情報も利用する. 本手法では, CiNii と国立国会図書館を対象としてメタ情報の取得を行う.

取得したメタ情報は, 本システムで使用する語彙と異なる場合には, 語彙あるいは情報の翻訳が必要となる. その際に本研究では, 語彙の意味が失われないよう, 語彙の定

表 3 対象が論文の場合の属性

名前空間	語彙	説明	
dc	title	タイトル	自動
dc	date	作成日	自動
dcterms	format	記録形式	自動
foaf	maker	著者情報	自動
foaf	Group	所属	自動
dcterms	description	内容記述	あれば自動
rdfs	seeAlso	情報提供元	あれば自動
dc	type	リソースタイプ	手動
dc	publisher	公開者	手動
dc	language	本文の言語コード	手動
dcterms	abstract	要約	手動
dcterms	accessRights	アクセス権	手動
dcterms	references	参考文献	手動
dcterms	subject	キーワード	手動

成果物の種類 論文

メタ情報を編集する(論文)

タイトル* 研究成果物のメタデータと管理データベース.pdf

著者情報
 作者名 西村 友基
 所属 A6研究室
 追加

公開者 公開者

公開日* 2017.06.29

言語コード* 日本語

要約 要約

参考文献 参考文献
 追加

キーワード キーワード
 追加

アクセス権* 公開

登録 クリア

図 2 情報入力画面

義を確認し、意味が同義または定義域の大きいものを用いるものとした。この翻訳部分を取得先毎に準備することで、他機関の RDF 情報取得も可能とした。

3.4.4 データ出力

RDF により表現したメタ情報は、間接的に参照できるようファイルで出力する。RDF データを記述する形式は多数存在しているが、同じ RDF データを別の形式で記述しても全く同じトリプルになりえ、論理的に同等である[3]。そのため、本システムでは W3C 標準である RDF/XML ファイル形式で出力する。したがって、RDF ファイルを参照することで、管理データベースから情報を取得した場合と同等の情報を取得できる。

4. 評価・考察

本手法のメタ情報入力とデータ出力について評価する。

研究成果物のメタ情報入力の評価として、Mahara を利用している大学生 9 人を対象として、本システムを利用後にアンケートを実施した。アンケート項目について 1~5 の 5 段階 (5 が最良) で評価を行った。アンケート項目と結果を表 4 に示す。この結果から、本システムの表記や操作は

表 4 評価アンケート

	内容	評価の平均
Q1	表記は分かりやすいか	4.3
Q2	入力操作は分かりやすいか	4.7
Q3	入力が手間でないか	4.4
Q4	入力項目が十分であるか	4.2

わかりやすく、入力に手間がかからないことが推察される。

管理データベースから出力された RDF ファイルに関して、正しく表現されているかを W3C が提供する RDF Validator を用いて 80 件 (CiNii より 20 件, 国立国会図書館より 20 件) 検証した結果、すべて正しく表現できていたため、他のソフトウェアからも参照可能である。

非公開の研究成果物に対し、自動または手動で取得した情報を RDF で表現し付与・管理することで、非公開の研究成果物の解析・検索に必要な情報を付与する仕組みを定型化した。これを利用することで、解析・検索にかかる手間を減らすことができると考える。

5. おわりに

本稿では、Mahara の研究成果物を対象としたメタ情報の管理手法について提案した。提案手法では、研究成果物にメタ情報を付与し管理する手法を提案した。メタ情報はメタ情報記述のための枠組みである RDF に基づき表現することで、意味に基づいた検索や、他の公開されているメタ情報との連携を可能とした。

課題として、今回 Mahara の研究成果物として対象とした論文、発表スライド、議事録以外にも対象となりうる成果物があるため、新たに対象となる成果物の属性の定義・登録とフォーム処理の追加を行い、他の成果物に対応することが挙げられる。

今後は、自動付与可能なメタ情報の拡張、メタ情報の関連性、類似性、時系列等から研究成果物間の関係性を抽出しデータ化・可視化・提示を行いたい。

謝辞

本研究は JSPS 科研費 15K12168 の助成を受けたものです。

参考文献

- [1] Mahara ePortfolio System
<https://mahara.org/>
- [2] 野村 卓哉, 佐野 雅彦, 松浦, 健二, 関 陽介, 上田 哲史, “Mahara におけるデータ間の関係性解析支援システム”, 電子情報通信学会総合大会講演論文集, Vol.2016, No.1 (2016).
- [3] RDF 1.1 Primer
<http://www.w3.org/TR/rdf11-primer/> (2014).
- [4] 廣川 佐千男, 酒井 敏彦, 曾 駿, 殷 成久, “クロス集計による文獻ファセット検索システムの提案”, 研究報告情報基礎とアクセス技術 (IFAT), Vol.2011, No.1 (2011).
- [5] NGUYEN MANH CUONG, 加藤 大智, 橋本 泰一, 横田 治夫, “論文のメタ情報を利用した研究履歴自動抽出・可視化システム”,
- [6] 槇 俊孝, 若原 俊彦, “論文メタデータ解析による日英論文の関連付けと時系列動向要因分析の検討”, 電子情報通信学会技術研究報告, Vol.113, No.479 (2014).
- [7] 森本 康彦, 喜久川 功, 宮寺 庸造, “e ポートフォリオ活用のための蓄積文法と支援システムの開発”, 日本教育学会論文誌, Vol.35, No.3 (2011).
- [8] 中村 祐基, 中島 裕聡, 高木 正則, 山田 敬三, 佐々木 淳, “e ポートフォリオからのベストワーク自動抽出手法の提案”, Vol.2013, No.1 (2013).
- [9] 大向 一輝, “学術情報サービスのメタデータ・デザイン”, 情報の科学と技術, Vol.60, No.12 (2010).
- [10] DBpedia
<http://wiki.dbpedia.org/>