

研究図書館の活動成果の関連付けによる 知識活用支援に関する研究

中村覚¹ 大和裕幸¹ 稗方和夫¹ 満行泰河¹ 牧野元紀² 會谷佳光²

Satoru Nakamura¹, Hiroyuki Yamato¹, Kazuo Hiekata¹, Taiga Mitsuyuki¹, Motonori Makino², and Yoshimitsu Aitani²

¹ 東京大学
¹The University of Tokyo

² 東洋文庫
²Toyo Bunko

Abstract: 研究図書館の役割は、主に「図書等の研究資料の管理」「研究者による研究活動」「展示による啓蒙活動」である。これらの活動は相互に依存しており、所蔵資料に基づく研究を行い、展示等を通じた資料や研究成果の公開を行う。しかし、これらは担当組織の違い等により、密に連携しているとは言いがたい。本研究では、これらの活動成果に関する情報を **Linked Data** を用いて関連づけ、研究図書館における知識活用を支援することを目的とする。

緒言

研究図書館とは、貴重書等の資料を収集・所蔵し、それら一次資料を用いた史料研究を行う機関である。具体的な活動内容は、「図書等の研究資料の管理」「研究者による研究活動」「展示による啓蒙活動」である。これらの活動は相互に依存しており、研究者は所蔵資料に基づく研究を行う。また、展示会の企画等を通じ、それら一次資料や研究成果を一般に公開する。しかし、これら活動はそれぞれ異なる専門性を必要とし、担当する組織が異なることが一般的である。この結果、組織間の活動やその成果の連携が困難である。この課題に対し、本研究では研究図書館における組織間の知識活用を支援することを目的とし、上述した各組織の活動成果である「研究資料」「研究論文」および「展示資料」を統合的に管理するための手法を提案する。具体的には **Linked Data** を用い、組織毎に異なる **DB** 上で管理・公開されている各種資料の関連付けを行う。また、研究図書館の一つである東洋文庫[1]を対象とし、本手法に基づいて関連づけた資料情報の活用方法について検討する。

展示企画プロセス

本研究では、研究図書館における知識活用の例として、展示企画プロセスを対象とする。ここでは展示企画プロセスについて概説し、課題とその解決に必要な要件について整理する。

展示会の企画を行うにあたり、まず展示会のテ

マ選定を行う。このとき、研究図書館が所蔵する貴重書や研究者による研究成果、過去の展示情報を参照する。また、展示企画の目的は主に啓蒙活動であり、多様な知識や興味を持つ来訪者に対して、展示資料の背景説明、および地図や年表等を用いた多角的な情報提供を行い、理解度向上を支援する。

このプロセスの支援には、まずテーマ選定時に参照する「研究資料」や「研究論文」、および過去の展示に関する「展示情報」を統合的に収集可能な環境が必要となる。しかし先述したように、これらの資料は異なる組織で管理されており、統合的な管理には **DB** の違いや記述形式のバラツキ等が課題となる。また、展示資料に対する来訪者の理解度向上を目的とし、各資料に対してキーワードや著者情報、および出版地に関する位置情報や歴史上の出来事に関する年表情報を関連づけて管理する必要がある。

本研究では上記に示した必要要件および課題に対して、書誌情報の標準化を行い、**Linked Data** を用いた各種 **DB** の関連付けを行う。また **DBpedia** 等の **LOD** を活用し、各資料の関連事項の整理を支援する。

関連研究

博物館や美術館、公文書館、図書館などでは、それらが保有する文化資源をデジタル化し、インターネット上で一般に公開するデジタルアーカイブの提供が進められている。デジタルアーカイブの利点として、文化資源のデジタル化による劣化防止、時間的・地理的制約を超えた資料提供、**DB** 化による新たな

情報発信等が挙げられる。また近年では、Linked Data を用いたデジタルアーカイブの統合が注目されている[2]。Linked Data とは、RDF (Resource Description Framework) 等の機械可読な形式で情報を記述し、データ共有を支援する方法である[3]。本手法の活用例として、Europeana[4]は図書館を始めとする文化施設が保有する文化資源の情報を RDF に変換し、ポータルサイトの提供による多数の文化資源へのアクセスを可能としている。また公開情報に対する検索を可能とする SPARQL Endpoint を提供し、分野やシステムを超えたコンテンツやメタデータの共有・利用を促進している[5]。国内における活用事例としては、国立国会図書館が情報資源の組織化および利用提供のためのメタデータ標準として「国立国会図書館ダブリンコアメタデータ記述 (DC-NDL)」を定めている[6]。この DC-NDL に基づき、国立国会図書館が所蔵する図書や論文等の書誌情報を RDF 形式で管理・公開している。

本研究ではこれら取り組みを参考とし、研究図書館において活動内容の異なる組織が管理する DB を Linked Data を用いて関連付ける。さらに、国際的なメタデータ標準である Dublin Core [7]、および DC-NDL 等を用いることで、組織毎に異なる記述方法で管理されている各種資料の書誌情報の標準化を行う。

提案手法

提案手法の概要を図1に示す。「研究資料」「研究論文」「展示資料」等の各種 DB の統合を目的とし、本手法は以下に示す2つのプロセスから構成される。

- ① 書誌情報の標準化と RDF データ化
- ② RDF データの関連付けと書誌情報の拡充

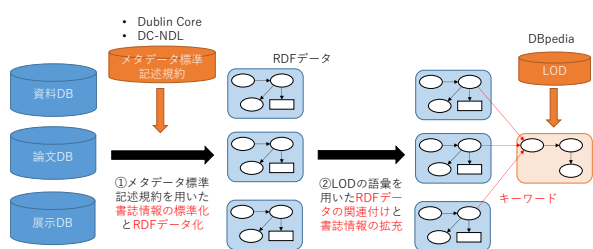


図1: 提案手法の概要

1点目のプロセスでは、各DBの書誌情報のRDFデータ化を行う。このRDFデータ化において、DB間の連携には記述形式の標準化が重要となる。本手法ではメタデータ記述形式として、Dublin CoreおよびDC-NDLを用いる。2点目のプロセスでは、DB毎に作成したRDFデータの関連付けを行う。各種資料の書誌情報からキーワードを抽出し、DBの関連付けを行う。加えて、対象キーワードをLODの一つで

あるDBpedia[8]と関連づけることにより、DBpediaが提供する情報に基づく書誌情報の拡充を行う。具体的には、対象キーワードの概要説明やサムネイル画像、地理情報等の利用が可能となり、これらの情報を用いた多角的な資料提供を行う。

以下、各プロセスの詳細について述べる。

書誌情報の標準化とRDFデータ化

まず「研究資料」「研究論文」「展示資料」の書誌情報をRDFデータ化する。このRDFのメタデータ記述形式として、Dublin CoreおよびDC-NDLを用いる。特にDC-NDLでは「図書」「論文」等の種別毎にメタデータ記述のためのRDFスキーマを提供している。表1に図書情報を記述するためのRDFスキーマ例を示す。「タイトル」や「著者」等を記述するための語彙に加え、DC-NDLによって拡張された「シリーズ名(業書名)」や「請求記号」等、図書に関するメタデータに適した語彙を提供する。これら種別毎に提供されるRDFスキーマを用い、研究図書館が保有する各種資料の書誌情報をRDFデータ化する。

表1: 図書情報のRDFスキーマ例

データ項目	語彙	例
タイトル	dcterms:title	中国の二つの悲劇: アヘン戦争と太平天国
著者	dc:creator	増井経夫 著
シリーズ名	dcndl:seriesTitle	研文選書; 1
出版地	dcterms:abstract	東京
出版社	dcterms:publisher	研文出版
出版年	dcterms:issued	1978
大きさ、容量等	dcterms:extent	247p; 20cm
請求記号	dcndl:pageRange	E-222.065-マ301-001

RDFデータの関連付けと書誌情報の拡充

次に前プロセスでDB毎に作成したRDFデータの関連付けを行う。この関連付けの方法を図2に示す。

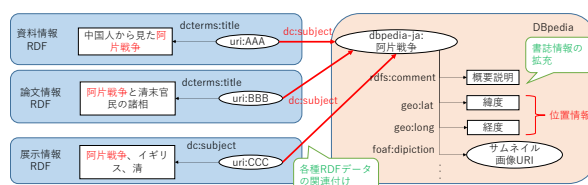


図2: RDFデータの関連付け

各種資料の書誌情報は、基本的に自由記述形式の文字列で記述される。この書誌情報内に含まれるキーワードを抽出し、主題を記述するための語彙である「dc:subject」を用いることで、各種資料のキーワードをResource形式で関連付ける。このResourceを表現するためのURIとして、DBpediaが定義する語彙を利用する。これにより、各種RDFデータの関連

付けに加え、DBpedia が提供する書誌情報に基づくメタデータの拡充が可能となる。

このプロセスにおいて、抽出対象とするキーワードの作成手順を図3に示す。

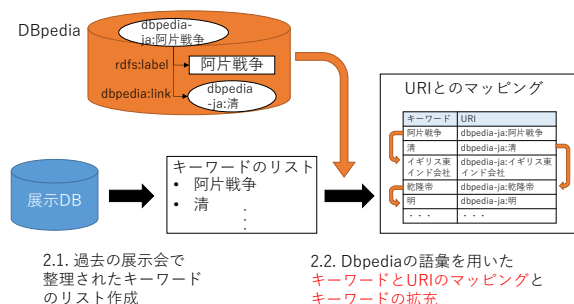


図3：キーワードのリストの作成手順

展示会はあるトピックに基づいて企画され、来訪者への情報提供を目的として、各展示資料のキーワードを整理するのが一般的である。このため、まず展示DBの情報を用い、過去の展示会で整理されたキーワードのリストを作成する。

次に、キーワードとURIのマッピング、および拡充を行う。DBpediaは、先述した各キーワードのURIに加え、ラベル名やWikipediaのページ内でリンクする他のキーワードのURIを提供している。この情報を利用して、先に展示DBから抽出した各キーワードに対してURIをマッピングし、さらにリンク関係にある他のキーワードのラベルとURIを追加する。

上記により作成したキーワードリストを用い、「資料」「論文」「展示」に関するRDFデータ、およびDBpediaを関連付ける。

適用事例

本適用事例では、東洋学の専門図書館ならびに研究所である東洋文庫が提供・管理を行うDBを対象とする。対象とする「図書DB」「論文DB」および「展示DB」の関係を図4に示す。



図4：対象とするDB

各DBはそれぞれ「図書部」「研究部」「普及展示

部」が異なるドメインで管理・運用しており、これら各組織の活動成果の分散が課題となっている。

本適用事例では提案手法に基づき、各種DBをLinked Data化して関連づける。また、特に各組織の活動成果を一般に公開する「普及展示部」の活動支援を目的とし、各種DBの情報を相互に利用し、所蔵資料を多角的に提供するシステムを開発する。以下では、関連づけた各種DBの活用例として、システムが提供する2つの機能について述べる。

開発したシステム

システム構成を図4に示す。開発したシステムはSPARQL Endpointを用い、提案手法に基づいて構築したRDFデータを利用する。なお、SPARQL Endpointとして、RDFストアとRDFに対する検索や登録等のAPIを提供する「KASHIWADE[9]」を利用した。

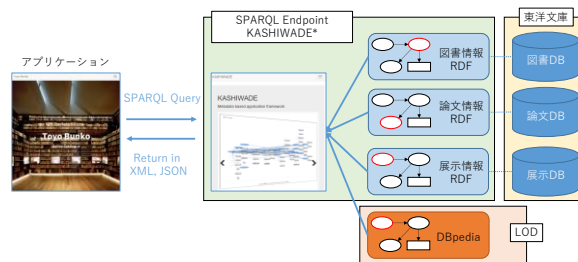


図5：開発したアプリケーション構成

機能①：多角的な資料提供

一般利用者への情報提供を行う展示会では、所蔵資料を様々な観点から提供する必要がある。多角的な資料提供の支援を目的としたユーザインタフェースを図6に示す。



図6：多角的な資料提供

本インタフェースは、資料の出版地に基づく位置情報、資料が掲載された展示会の図録情報、資料の著者に基づく人物情報、資料の出版年に基づく年表情報など、多角的な観点から資料検索を可能とする。これらで必要とする位置情報やサムネイル画像は、各資料に付与されたキーワードに基づき、DBpedia

が提供する情報を利用している。

機能②：DB の関連付けによる資料管理

各種異なる DB で公開されている資料の統合管理を目的としたユーザインタフェースを図 7 に示す。各種資料に付与されたキーワードに基づき、「図書」「論文」「展示」に関する書誌情報と、DBpedia が提供する情報を一つのインタフェースで管理する。各資料の書誌情報には、RDF データの作成元となる DB へのリンクが付与されているため、画像資料や公開済み論文、詳細な書誌情報等を各種 DB で確認することができる。



図 7：各種資料の統合管理

考察

書誌情報の標準化

「研究資料」「研究論文」「展示資料」等の各種 DB の関連付けを目的とし、Dublin Core および DC-NDL をメタデータの記述形式として利用し、書誌情報の標準化を行った。これにより、各種 DB の書誌情報を一つのインタフェースで提供する統合管理が可能となった。また、特に東洋文庫における「資料 DB」については、各研究組織およびプロジェクト毎に DB を公開しており、「資料 DB」内でも資料の分散が課題であった。これに対しても書誌情報の標準化により、資料情報の統合検索が可能となった。今後は標準化された書誌情報の活用に向け、東洋文庫が所蔵する研究資料だけでなく、国立国会図書館等が LOD として公開している各種資料との関連付けを行う。

研究図書館における活動成果の統合

研究図書館の活動成果の統合、および知識活用支援を目的とし、所蔵資料を多角的な観点から提供するためのシステムを開発した。このような多角的な情報提供を行うにあたり、DBpedia を例とした LOD と各種資料情報を関連付けることにより、位置情報や人物情報、年表情報に基づく資料検索が可能となった。これにより、所蔵資料に対する知識を持たな

い一般利用者に対し、背景知識や関連事項を組み合わせた資料提供を行うことで、所蔵資料および研究成果の公開による一般利用者への啓蒙活動を支援することができると思う。

結論

本研究では研究図書館における組織間の知識活用を支援することを目的とし、各組織の活動成果である「研究資料」「研究論文」および「展示資料」を統合的に管理するための手法を提案した。また、知識活用の例として展示企画プロセスを対象とし、各種 DB の書誌情報の標準化による統合管理、および LOD との関連づけによる多角的な資料提供を可能とするシステムを開発した。

今後は開発したシステムの有用性の検証を目的とし、研究図書館の活動における実利用を通じた評価を行う。具体的には、展示会におけるテーマや資料選定、およびキャプションの整理等の展示企画活動プロセスの効率化等について検討する。

参考文献

- [1] 東洋文庫, <http://www.toyo-bunko.or.jp/>, (accessed 2016-02-22).
- [2] 橋詰秋子: なぜ図書館は Linked Data に取り組むのか 欧米の事例から, 情報管理, Vol.58, No.2, pp.127-134(2015).
- [3] 武田英明: 動向レビュー: Linked Data の動向, カレントアウェアネス, No. 308, pp.8-11(2011).
- [4] Europeana, <http://www.europeana.eu>, (accessed 2015-11-06).
- [5] Europeana Creative Challenges, <http://pro.europeana.eu/web/europeana-creative/challenges>, (accessed 2016-02-22).
- [6] 国立国会図書館, http://www.ndl.go.jp/jp/aboutus/standards/meta/dcndl_examples.html, (accessed 2016-02-22).
- [7] ダブリンコア (Dublin Core), <http://www.ndl.go.jp/jp/aboutus/standards/translation/dcmi-terms.htm>, (accessed 2016-02-22).
- [8] DBpedia Japanese, <http://ja.dbpedia.org/>, (accessed 2016-02-22).
- [9] 中村覚, 大和裕幸, 稗方和夫, 岡田伊策, 齋藤稔, 笈田佳彰. ドメイン知識の記述支援と活用のためのデジタル資産管理システム基盤「KASHWADE」の開発 - 情報システム開発における設計書・設計知識への適用, 2015 年度人工知能学会全国大会, 2M5-4, 2015.