

図書館情報学アーカイブズ理解を目的とした DBpedia の活用

渡部航太郎^{†1} 松村敦^{†2} 宇陀則彦^{†2}

概要: 近年、デジタルアーカイブの増加に伴い、一般の人がアーカイブ資料に触れる機会が増えた。しかしながら、利用者の普及ができたとは言えない。原因として、アーカイブ資料を理解するための知識が少ないからと考えられる。したがって、本研究ではアーカイブ資料と DBpedia を繋げたデジタルアーカイブを構築し、利用者がアーカイブ資料に関する知識を随時得られるようにした。これにより、利用者はアーカイブ資料に関わる情報を補充しながら閲覧をすることができる。その結果、利用者はアーカイブ資料の理解促進をすることが可能となった。

キーワード: デジタルアーカイブ, Linked Open Data (LOD), DBpedia

Application of DBpedia for Understanding Archives of Library and Information Science

KOTARO WATANABE^{†1} ATSUSHI MATSUMURA^{†2}
NORHIKO UDA^{†2}

Abstract: In recent years, with the increase in digital archives, opportunities for ordinary people to touch archive materials have increased. However, it cannot be said that users have spread. It seems to be because the knowledge for understanding the archive materials is small as a cause. Therefore, in this research, we built a digital archive linking archived materials and DBpedia so that users can obtain knowledge related to archive materials from time to time. As a result, the user can browse while complementing the information related to the archive material. As a result, it became possible for users to promote understanding of archive materials.

Keywords: Digital Archives, LOD, DBpedia

1. はじめに

文書館の管理に関する法律の第二十三条には「国立公文書館等の長は、特定歴史公文書等について、展示その他の方法により積極的に一般の利用に供するように努めなければならない」との記述がある。この条文より、文書館は本来、一般の人を広く利用へ導かなければならないが、それを十分に果たせているとは言えない。資料を一般の利用へ導く方法としてデジタルアーカイブがある。デジタルアーカイブは近年増加し、広くアクセスできるようになったが、アーカイブ資料を単純にデジタル化しただけという場合が多く、利用が広がったとは言えない。その原因として考えられることに、一般の利用者はアーカイブに対する知識が少なく、理解することが難しいからだと考えられる。そこで本稿では、アーカイブ資料の情報を他の情報資源と結びつけることによって、アーカイブ資料に対する知識が少ない人に対してアーカイブ資料への理解促進を図ることを目的とする。

2. Linked Open Data

情報を繋げる手法として近年注目を集めているのが Linked Open Data(以下、LOD と称す)である。LOD とは Resource Description Framework(以下、RDF と称す)によって、データに意味を持たせることとデータを繋げることを可能にしている仕組みである。LOD には様々な種類がある。具体例として LODAC プロジェクト[1]と、Web NDL Authorities[2]と、DBpedia[3][4]というプロジェクトをあげる。

LODAC プロジェクトは、博物館・美術館、生物種に関わる国内の情報資源を中心に多くの人々に有用な学術情報を Linked Data として公開し、それらを利用したアプリケーション等を開発しているプロジェクトである。LODAC Museum として web 上にデータを公開している。2012年6月6日現在で収蔵品全体のデータは約177万件あり、作者に関しては約8800件、施設は約20万件存在している。Web NDL Authorities とは、「国立国会図書館典拠データ検索・提供サービス」のことで、国立国会図書館が維持管理する典拠データを一元的に検索・提供するサービスである。web ペ

^{†1} 筑波大学情報学群知識情報・図書館学類
College of Knowledge and Library Science, School of Informatics, University of Tsukuba.

^{†2} 筑波大学図書館情報メディア系
Faculty of library, Information and Media Science, University of Tsukuba.

ージにて SPARQL 検索も可能であり、外部サイトとして Wikipedia のページのリンクがあることや、それぞれのデータに対して、RDF/XML 形式、RDF/Turtle 形式、JSON-LD 形式をえらぶことができる。DBpedia では、インターネット上にあるフリーの百科事典である Wikipedia から構造化データセットを抽出し、RDF を用いて記述している。DBpedia は主に英語版の Wikipedia をもとに作成されている。日本語版の DBpedia としては、DBpedia Japanese が存在している。

3. 図書館情報学アーカイブズ

本稿で用いるアーカイブズ資料は筑波大学図書館情報メディア系のプロジェクト[5]である「21 世紀図書館情報専門職養成研究基盤アーカイブ構築：図書館情報専門職の再検討」で扱った資料(以下、図書館情報学アーカイブズ称す)である。このプロジェクトでは日本の図書館専門職養成の中核的機関の一つであった図書館情報大学およびその前身校の約 1 世紀にわたる図書館専門職養成の沿革にかかわる歴史的資料を対象にアーカイブを構築した。

前身校アーカイブ資料文献資料と現物資料を調査すると前者は 2730 点、後者は 273 点あることを明らかにし、それぞれをアーカイブ化した。文献資料には以下のような資料が含まれていた。文書類（教育・研究資料、授業教材、パンフレット、附属施設資料、名簿、私信、手稿等）、写真資料（アルバム、ネガ等）、音声資料（講義音声、行事音声等）、映像資料（行事動画等）。これらの中で、著作権上問題のない資料について部分的に電子化を行なった。さらに図書館情報大学名誉教授佐藤隆司氏から寄贈された図書館短期大学名誉教授馬場重徳氏の個人文書をアーカイブに統合し概要目録を作成した。

現物資料は、図書館情報専門職養成に関わる用具（図書館業務用品、製本用具、資料修復用具等）、図書館情報専門職養成に関わる機器（情報検索機器、計算機、コンピュータ等）、図書館業務に関わる機器類（図書館家具、図書館機器、図書館業務用品等）があるが、これらすべてを電子化し中間目録を作成した。

本研究では、資料 ID「書架下-1-a-131」、タイトル「飛翔 - 新しい情報社会に向けて-」を選んだ。この資料は、外部へ向けた図書館情報大学の見学会時に作成された資料であり、画像資料とテキスト資料がバランスよく配置されていたためこの資料をシステム作成の際に使用した。

4. アーカイブズと DBpedia との接続

DBpedia は Wikipedia の情報を抽出化していることから、様々な分野の用語が集まっている。そのため、アーカイブ資料理解のための補助システムとして利用する際に大きく効果を発揮する。また、RDF 記述によって用語の説明部や他のページへのリンク部などの分類がなされているので、

情報の選択がやりやすいというメリットがある。

よって、アーカイブ資料に対して BDpedia Japanese を用いて他の情報と結びつける。結びつける語句は、「飛翔」、「1982年」、「図書館情報大学」(2カ所)、「クロック」、「Pascal」、「COBOL」、「ニューメディア」、「古典言語」、「シソーラス」、「英国放送協会」、「HITAC」、「PDP」、「エンドユーザー」、「類書」の計 14 種、15 カ所にアノテーションを設定した。アーカイブ資料と DBpedia の関係性は以下の図 1 に示す。

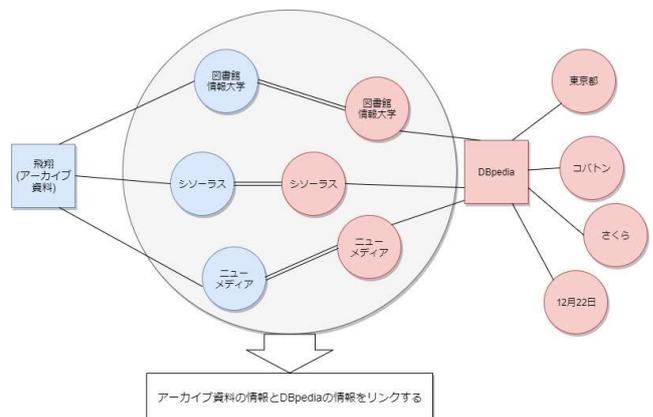


図 1 用語の繋がりを表すイメージ図

Figure 1 Connection of Terms

次にインターフェースの説明をする。

図 2 はアーカイブ資料を閲覧する画面となっている。画面左にアーカイブ資料の画像を載せ、画面右に用語検索をした際に表示することができるようになっている。機能として以下のものがある。

1. 資料の画像
2. 画像中の用語検索な箇所にあるマーカー
3. ページ送りのためのボタン
4. 用語検索にて調べた用語の一覧を示すためのボタン
5. 4 番のボタンを押した際に表示されるリスト

1 と 3 に関しては、画像の表示をしているもの、資料のページを送ったり戻ったりするボタンである。2 のマーカーに関しては資料中にて、アーカイブ資料を閲覧する際にわからなると予想される用語や画像に対してこちらでマーキングしたもとなる。こちらをクリックすることでクリックした対象を調べる用語検索画面へと移動する。4 と 5 は用語検索画面にて調べた用語を列挙するための機能で、2 のマーカーをクリックするたび更新されるようになっているので、検索用語ごとに用語検索の関係性を可視化することができる。

図 3 は図 2 で選択したアノテーションに関係する用語を DBpedia に接続して検索を行った結果を表示する画面である。この画面では以下の機能がある。

1. 検索用語
2. 用語の説明文

3. 関連用語
4. 関連用語の更新ボタン
5. 用語ページの戻るボタン
6. アーカイブ資料画像に戻るリンク

1 の検索用語が DBpedia を検索する際の Subject となり、2 の DBpedia で検索した用語の Predicate のなかで「<http://www.w3.org/2000/01/rdf-schema#comment>」となっている Object を表示した。3 の関連用語は 1 で検索した用語で Predicate が「<http://dbpedia.org/ontology/wikiPageWikiLink>」となっている Object を表示させる。該当する Object が 10 個以下の場合には全て表示し、10 個より多い場合、その中からランダムに 10 個表示する。3 の用語をクリックした際には、クリックした用語に対しても同様に検索と結果表示がなされる。4 に関しては関連用語から別の用語を検索したいが、その 10 件の中には検索したい用語がない場合、別の 10 件を表示させるために用いる。5 番は、用語の検索時に別の用語を調べたいが、1 つ前の用語に戻って続けたい場合に用いる。6 は用語検索を十分に終了した後、アーカイブ資料画面に戻りたい場合クリックするリンクである。

これら機能を実装することによって、アーカイブ資料と資料中の用語の関係性、資料中の用語とその用語と関係性のある用語の可視化をすることができる。

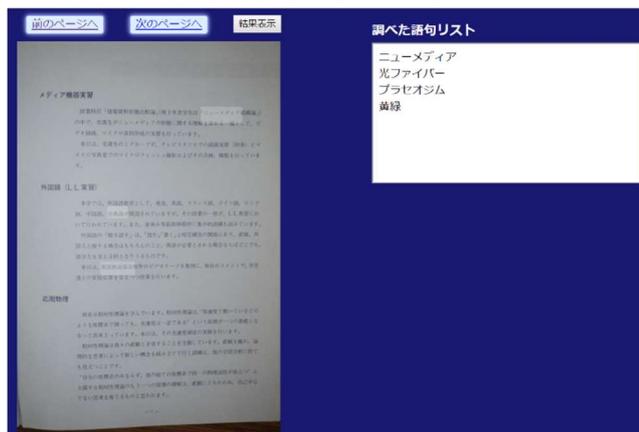


図 2 アーカイブ資料表示画面
Figure 2 Display of Archives Documents

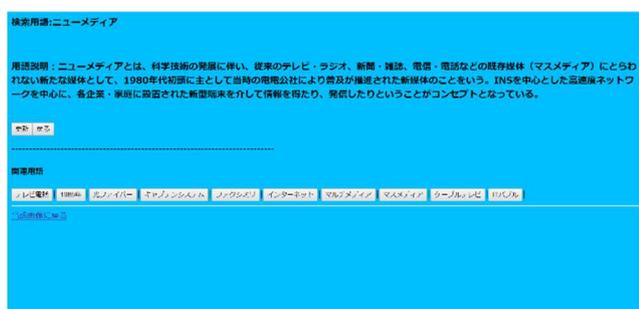


図 3 用語検索画面
Figure 3 Display of Retrieval

5. 実験概要

利用者がアーカイブ資料の理解促進をすることができたのか確認するための評価実験では、アーカイブに対して知識を有していない大学生 10 名に対して行った。実験方法では、アーカイブの画像のみのシステムを使用した非利用群と DBpedia とアーカイブを繋ぎ合わせたシステムを使用した利用群の 2 つに分けた。評価方法として、作成したデジタルアーカイブに関する質問紙調査を行った。質問リストを表 1 に示す。

表 1 質問紙の設問一覧

Table 1 Questions about LIS Archives

問 1	図書館情報大学は何年に設置されたものですか
問 2	図書館情報大学の同窓会の名称は
問 3	HITAC とはどこの会社が使用していた商標か
問 4	古典語とは何に使われた言語か
問 5	古典語には具体的にどういった言語があるか
問 6	ニューメディアとはどういった媒体か
問 7	ニューメディアの具体的な例として何が挙げられるか
問 8	エンドユーザーとは何を意味する言葉か

分析の基準として、実験参加者が図書館情報学アーカイブの内容について理解しているかを検証するため作成した調査用紙を用いて分析を行った。その際、記述内容がどのような内容なのかを分析の基準とした。そして、回答欄に記入した内容がシステムにて補完された知識であるかを判断した。

6. 実験結果

実験結果は表 2 に示す。被験者 ID が A,B,C,D,E が非利用群で、被験者 ID が F,G,H,I,J が利用群である。1 問 1 点である。システム利用群では最高得点が 6.5 点、最低得点が 2.5 点、平均点は 4.5 点だった。非利用群では、最高得点は 2 点、最低得点は 1 点、平均点は 1.4 点だった。問題別では、問 3 の 8 人が最も正解数が多く、次に多いもので、問 8 の 7 人だった。また、問 5 が 1 人部分点のみで最も正解数が少ない問題となった。次に少ない問題として、問 7 の 1 人部分点で、1 人正解したという結果になった。

問 1 は 1961 年から 1988 年と幅広い年代が答えに上がった。1 人は筑波大学と統合した年と閉校した年を答えていた。問 2 は「橋会」、「茗溪会」、「ULIS」、「SLIS」といった回答があがった。問 3 は、「日立(製鉄所)」という答えでほとんど一貫していた。問 4 は、「古典文学」という答えのほか、プログラミング言語のものと勘違いしている答えが多くみられた。問 5 は、問 4 に引き続きプログラミング言語を答えに出している人が多くあげられた。問 6 は正解者以外には「新しいメディア」と答える人が多く、直訳しただけの答えが見られた。問 7 は、「パソコン」、「テレビ」、「ラ

ジオ」といった答えや「フロッピーディスク」や「マイクロフィルム」を答えにあげる人もいた。問8では、「商品の利用者」に加え、「最終利用者」と直訳する人が多くあげられた。

表2 被験者の正誤表
 Table 2 Subject's answer table

被験者 ID	問題								合計点 (点)
	1	2	3	4	5	6	7	8	
A	×	×	○	×	×	×	×	○	2
B	×	○	×	×	×	×	×	○	2
C	×	○	×	×	×	×	×	×	1
D	×	×	○	×	×	×	×	×	1
E	×	×	○	×	×	×	△	×	1.5
F	×	×	○	×	×	△	×	○	2.5
G	○	○	○	○	△	○	×	○	6.5
H	×	×	○	○	×	△	○	○	4.5
I	×	○	○	×	×	○	×	○	4
J	○	×	○	○	×	○	×	○	5
正答数 (人)	2	4	8	3	0.5	4	1.5	7	30

7. 考察

システムの利用群と非利用群の平均点の差は3点であり、本システムはアーカイブ理解の促進のための知識の補強に繋がった。特に、問6でその差が顕著に表れ、非利用者群は正答者がいなかったのに対して、利用者群は3人正解、2人が部分正解となった。また、問8では非利用者群が2人正解、利用者群が全員正解という結果から資料中の用語の意味を把握することに本システムは向いていることが分かった。一方、問5は非利用群が0人、利用群が1人正解であり、問7は非利用群が1人部分点で、利用群で1人正解という結果になった。この2問は共に用語の具体例を問う問題であった。これらの問題で正答するためには、用語検索の中で関連用語を見る必要があり、本システムの利用者はそこまで見ていなかった、または、利用したが知識として把握できなかったとわかる。このことから本システムでは、具体例を知ることまでは繋がらなかったということがわかる。

ここで注意しなければならないのは「知識」といった場合には2つの側面があることである。一つは人の頭の中にある知識、もう一つはアーカイブズや Wikipedia などの記録された知識である。図書館情報学アーカイブズの知識と DBpedia の知識が接続することで知識が増えるという現象は、記録された知識が増える（接続することで拡大する）

という意味であり、それを読んだ人の知識が増えるという現象とは区別される。さらに複雑なのは記録された知識と人の頭の中の知識には相互作用がある。本研究はこれらの現象の違いを意識して読み解く必要がある。

別の観点として、記録された知識をどう接続するかという問題がある。国立歴史民俗博物館が進めている総合資料学は考古学、歴史学、地理学、分析化学、文化財科学の資料を相互に接続し、異分野を横断した研究基盤を構築することを目指している。必然的に接続の際に要求されるのは学術的な正確性である。本研究は一般の人を広く利用へ導くことを研究動機としているため、接続における正確性は重視していないが、図書館情報学アーカイブズもアーカイブズである以上、信頼度の高い知識を接続すべきである。

8. おわりに

本研究では、アーカイブ資料理解のためにアーカイブ資料と DBpedia を繋げるシステムを作成し、被験者実験を行った。結果として、アーカイブ理解のために本システムは有効だとわかった。しかし、DBpedia への接続のみを行ったことや、本システムを利用しても正解しなかった問題があることが挙げられる。

今後の課題として、DBpedia 以外の LOD と接続し、検索可能なデータの拡張を図るとともに、知識の補助をより具体的な用語へも可能にしたいと考える。

謝辞

本研究は JSPS 科研費 16H02913 の助成を受けて行われた。

参考文献

1. LODAC. “LODAC PROJECT”. LODAC. <http://lod.ac/>, (2017-12-20).
2. 国立国会図書館. “Web NDL Authorities 国立国会図書館典拠データ検索・提供サービス”. Web NDL Authorities. <https://id.ndl.go.jp/auth/ndla>, (2017-12-18).
3. DBpedia Japanese. “DBpedia Japanese”. DBpedia Japanese. <http://ja.dbpedia.org/>, (2017-12-18).
4. 加藤文彦. DBpedia の現在リンクトデータ・プロジェクト. 情報管理. 2017年, vol. 60, no. 5, p.307-315.
5. 水嶋英治. 21世紀図書館情報専門職養成研究基盤アーカイブ構築：図書館情報専門職の再検討. 2017年, p. 1-139.