

Linked Open Data を用いた地域史資料からの発見を促す 写真検索システムの構築

高橋 正輝[†] 奥野 拓[‡] 川嶋 稔夫[‡]

公立はこだて未来大学大学院[†] 公立はこだて未来大学[‡]

1 はじめに

近年、文書や古写真、古地図などの地域史資料のデジタル化が広く進められている。それらのデジタル化した地域史資料と目録（メタデータ）を併せて保存・蓄積することは、デジタルアーカイブと呼ばれている。筆者らの所属する公立はこだて未来大学では、函館市中央図書館と連携し、函館圏における地域史資料のデジタルアーカイブの構築を進めている[1]。

函館の地域史資料の中には、写真が多く存在する。それは幕末に写真技術が導入されたことで、昭和前半までの多くの写真が公的に収集されたためである。しかし、その膨大さ故に、写真資料のメタデータを十分に整備することができていない。これらの写真を活用していくためには、限られた人材やコストで、写真をいかにして編纂し、検索可能とするかが課題となる。

本研究では、標準的な形式である RDF を採用する LOD (Linked Open Data) の技術を用いる[2]。写真とメタデータだけでなく、文献資料である函館市の年表を LOD 化することによって、様々な歴史的観点で写真を関連付けることが可能となる。これらの LOD を用いてユーザが求める写真だけでなく、歴史的な発見のあるような関連写真の検索が可能なシステムの構築を目指す。

2 LOD による写真の関連付け

Web においてデータを共有する取り組みとして近年注目を集めている LOD がある。LOD では、データ形式として RDF を採用していることから、リソース（事物）に意味を持たせたリンクを張ることでリソース間の関係を一意に表現することができる。また、RDF データは外部から参照可能な形式で公開する。そのようにすることで、同様の方法で公開されている様々なリソースを SPARQL というクエリ言語を用いて意味的関連で関連付けて取得することを可能にする。

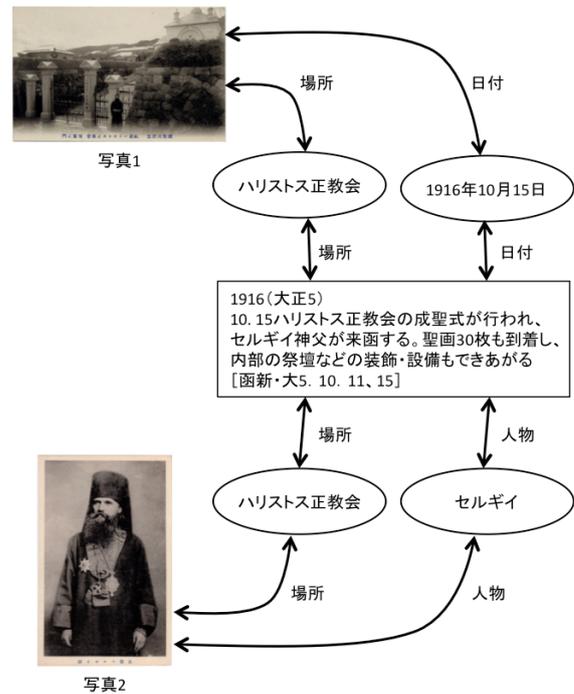


図 1 日付・人名・場所による写真の関連付け

写真とメタデータ、文献資料である「函館市史年表編」を LOD で公開することで、歴史的関連性のある写真が関連付く可能性がある。

「函館市史」[3]とは、函館の歴史をまとめた全 11 巻の書籍である。その中でも年表編は、旧石器時代後期から 2004 年(平成 16 年)まで記述された函館の歴史年表であり、PDF 化されている。内容については、主に明治以降が新聞資料を基礎資料として作成されており、西暦・和暦、月日とそれに対応した函館に関連する出来事が 700 ページにわたって記述されている。この年表を利用することで図 1 のように写真が関連付く可能性がある。図 1 の写真 1, 2 はデジタル資料館[4]で公開されている二枚の写真である。写真 1 のメタデータには、写真に関する出来事が起きた日付 (1916 年 10 月 15 日) と場所 (ハリストス正教会) が含まれている。1916 年 10 月 15 日の年表の出来事には、ハリストス正教会にセルゲイ神父が訪れた内容が記述されている。写真 2

Building a Photo Search System for Promoting Discovery from Historical Records with Linked Open Data

[†] Masaki Takahashi, Graduate School of Future University Hakodate

[‡] Taku Okuno, Toshio Kawashima, Future University Hakodate

表 1 日付・人物・場所を含む写真と年表の件数

	写真 (件)	年表 (件)
全件	13667	6808
日付を含む	383	6808
人物を含む	461	338
場所を含む	805	550

のメタデータには、年表に記述されていた人物（セルギイ）と場所（ハリストス正教会）が含まれている。従って、写真 1 と年表が日付と場所によって結びつき、年表と写真 2 が人名と場所によって結びつくことで、写真 1 と写真 2 が関連付く。このように、写真と年表に含まれる日付、人物、場所をリソースとして抽出し、RDF データを作成することにより歴史的な関連のある写真が関連付く可能性がある。本研究では、ユーザが求める写真だけでなく、日付、人物、場所の各々の観点で関連のある写真を検索できるシステムを構築する。そのシステムによって、ユーザに歴史的な発見を促すことを目指す。

3 RDF データの作成

写真のメタデータと年表から日付、人物、場所を抽出し、それぞれをリソースとする RDF データの作成を行った。日付はテキスト処理によって抽出した。人物については、「はこだて人物誌」[5] に掲載されている函館にゆかりのある人物 271 名と文字列マッチすることで抽出した。また、場所は、「函館市公式観光情報サイトはこぶら」[6] の歴史的建造物などを掲載している「見る」カテゴリの 198 スポットと文字列マッチすることで抽出した。表 1 は、写真と年表全件と日付、人物、場所を含むものの件数を示している。

RDF データは一般的に利用されている語彙を用いて作成を行った。写真のタイトルや内容説明、年表の出来事の説明には Dublin Core、人物には FOAF、場所には Schema.org を採用している。

4 検索システムの構築

作成した RDF データを用いた写真検索システムを構築する。写真と年表における日付、人物、場所のリソースを辿る方法は複数ある。あらゆる目的で写真を検索するユーザに対し、歴史的な発見を促すためには、様々なリソースの辿り方を利用した写真の検索ができる必要がある。

今後は、検索インターフェースについて検討を進める。

5 写真検索システムの評価

写真検索システムがユーザの歴史的な発見を促すことが可能か評価するための実験を行う。

実験方法としては、ユーザに本システムとデジタル資料館を使って調べたい写真を検索してもらい、その結果どちらが発見的な写真に辿り着くことができたか、本システムによって検索できる写真が発見的であると感じたかアンケート調査を行うことを検討している。

6 おわりに

函館の地域史資料である写真のメタデータと文献資料である年表の LOD 化による編纂と LOD を用いた写真検索システムについて報告した。写真のメタデータと年表に含まれる日付と人名、場所をリソースとした RDF データを作成することで、歴史的な関連性を持つ写真が関連付く可能性が高い。

今後は、リソースをいかに辿ることでユーザの発見を促すことが可能か検討する。その後、検索システムを構築し、ユーザによる評価実験を行う。

謝辞

本研究にあたってご協力を頂いた函館市中央図書館に感謝致します。

参考文献

- [1] 出口貴也, 中原裕成, 高橋正輝, 奥野拓, 川島捻夫: 地域の記録と市民の記憶を共有するデジタルアーカイブCMS, 第84回デジタルドキュメント研究会, 2011.
- [2] Tom Heath, Christian Bizer, Linked Data: Evolving the Web into a Global Data Space, Morgan Claypool Publishers, 2011.
- [3] 『函館市史』一覧, http://www.city.hakodate.hokkaido.jp/soumu/hen-san/shishi/shishi_ichiran.htm
- [4] デジタル資料館, <http://archives.c.fun.ac.jp/>
- [5] はこだて人物誌, http://www.city.hakodate.hokkaido.jp/soumu/hen-san/jimbutsu_ver1.0/index.htm
- [6] 函館市公式観光情報サイトはこぶら, <http://www.hakobura.jp/>