Vol.2013-DD-88 No.9
2013/1/18

函館の歴史資料を用いた地域写真アーカイブの編纂

高橋正輝[†1]　奥野拓[†2]　川嶋稔夫[†2]

函館のデジタルアーカイブに含まれる大量の写真資料はメタデータが少なく，写真間が歴史的関連で関連付けられていない．そこで，函館の歴史に関する文献を利用することで写真資料のメタデータを補い写真間が関連付く可能性がある．写真資料と函館市史年表編，はこだて人物誌を Linked Open Data (LOD)として作成し公開する．LOD とは外部とのデータ連携を実現する技術である．函館の歴史資料を LOD 化することで，地域写真アーカイブの編纂を目指す．

Compilation of Photo Archives Using Historical Records of Hakodate

MASAKI TAKAHASHI[†1]　TAKU OKUNO[†2]
TOSHIO KAWASHIMA[†2]

Photos in Hakodate photo archives have historical relations to each other, but there is no links between them. There is a possibility that the photos are linked by using historical records of Hakodate. This study aims to link the photos in the archives by generating Linked Open Data of photo archives, historical calendar, and historical figure of Hakodate.

1. はじめに

1.1 函館における地域アーカイブの取り組み

近年，文書や古写真，古地図などの歴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歴史資料と目録(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化と呼ばれている．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，函館圏における歴史資料のデジタルアーカイブの構築を進めている[1].

1.2 写真による地域史の編纂

函館の歴史資料の中には，写真資料が多く存在する．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．それらの写真資料は様々な年代の人物や建造物などの被写体を含んでおり，被写体間には歴史的な関連がある可能性が高い．地域の変遷を記録するためには，大量の写真資料を関連付けていくことが必要である．しかし，写真資料にはメタデータが十分に整備されていないため写真間の関連性を見出すことは難しい．また，公的に収集された写真(写真資料)は広報用の写真に限定されているため質的に薄いコレクションとなってしまっている[2].

1.3 歴史資料と市民の知識による地域写真の関連付け

函館には歴史に関する様々な文献が存在する．それらの文献は所有者や作成者が異なるが写真資料に関する情報を含んでいる．そのような散在している情報源を共有する技術として，近年注目を集めている Linked Open Data (LOD)がある．LOD では，データを RDF 形式で作成する．RDF を用いて，データ間に意味を持たせたリンクを張ることでデータ同士の関係を一意に表現できる．また，RDF を SPARQL のようなクエリ言語で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分野のデータを意味関係で関連付けて取得することが可能となる. LOD の形式で函館の歴史に関する文献と写真に付加されているメタデータを公開することで，歴史的関連性のある写真が関連付く可能性がある．また，函館に留まらず各地域の持つ歴史資料と関連付く可能性も考えられる．

一方で，市民の所有する写真は私的に撮影したものが大半であるが，地域の記録として価値ある事物や風俗が記録されている．そのため，その潜在的価値は極めて大きいと考えられる．さらに，メタデータが少ない写真資料に関しても，市民が被写体に関する情報を持っている可能性が高い．地域史の編纂を行うためには，写真資料だけでなく市民の所有する私的写真を含む地域写真を収集することも必要である．またそれらを整理する上で市民の参加は必要不可欠である[3].

本稿では，2，3 節において函館の歴史資料の LOD 化による写真資料の編纂について検討した結果について報告する．また 4，5，6 節では市民の知識を利用した関連付け手

†1 公立はこだて未来大学大学院
Graduate School of Future University Hakodate
†2 公立はこだて未来大学
Future University Hakodate

□

１９１２（明治４５）

２．１６　ロシア正教ニコライ大主教が東京神田駿河台のニコライ堂で死去する。２２日の葬儀にはニコライの後任者セルギイ大主教が祭主を務める［函日・明４５．２．２０、２４］

３．－　函館管内大沼における本年の採氷予定は約８０００トンと見込んで当業者と鉄道院との間で８０００トンの氷輸送を特約していたが、意外にも採氷成績良好により約１万トンに達する［函日・明４５．３．７］

－．－　この年の函館の戸数２万１２０４戸、人口９万４４６４人（男５万２９６８人、女４万１４９６人）となる［大１・区事務報告］

図 1 函館市史年表編の内容の例

法の提案とその評価結果について報告する．

2. 歴史資料を用いた地域写真の編纂

2.1 デジタル資料館における写真資料

函館市の歴史資料は函館市中央図書館が大量に所蔵している．それらの歴史資料のデジタルアーカイブは，デジタル資料館[4]で公開されており，写真資料については約 1 万 3 千点を閲覧できる．写真には資料番号が付加されており共通の資料番号を持つ写真は資料閲覧ページで関連資料として表示され，関連資料にアクセスすることもできる．しかしこれらの関連資料は，あらかじめまとめられた資料であり，歴史的な関連によって関連付けられている資料ではない．これらの大量の写真を歴史的に関連付けることで，歴史を辿った写真の検索が可能となる．そこで，写真間の歴史的関連を見出すために，函館市史年表編，はこだて人物誌の利用について検討した．

2.2 写真と函館市史年表編の結びつき

函館市史とは，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の書籍であり，PDF にもなっている．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て函館に関連する出来事が約 700 ページにわたって記述されている(図 1)．この年表を利用することで図 2 のように写真間を関連付く可能性がある．

図 2 の画像 1, 2 はデジタル資料館で公開されている二枚の写真である．画像 1 には，メタデータとして写真に関する出来事が起きた日付(1916 年 10 月 15 日)が含まれている．

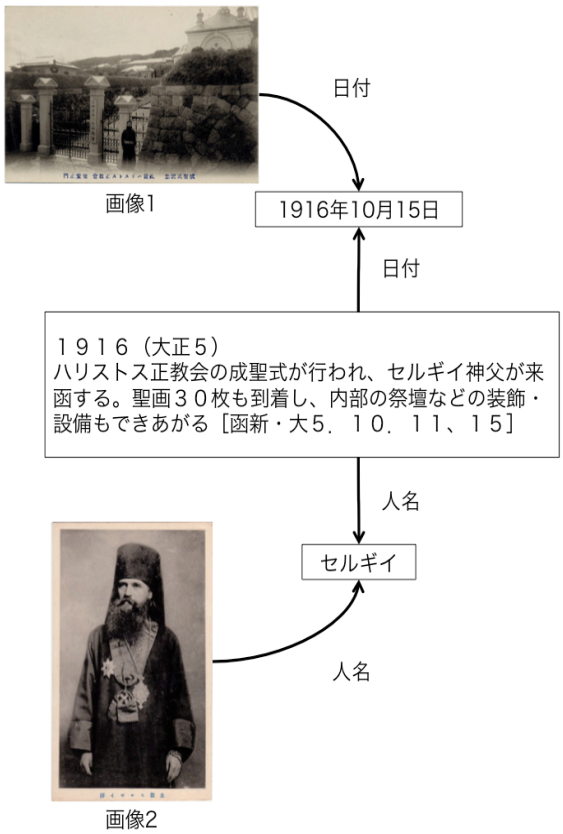

図 2　年月日と人名による結びつき

また，画像 1 には人物が写っているが人物についてのメタデータは含まれていない．画像 1 の日付を年表で調べたところ 1916 年 10 月 15 日にセルギイ神父が来函した内容が記述されていた．画像 2 には，その人名(セルギイ)がメタデータに含まれている．従って，画像 1 と年表が年月日によって結びつき，年表と画像 2 が人名によって結びつくことで，画像 1 と画像 2 の各々に写っている人物は同一の人物である可能性がある．このように，各々の資料における年月日や人名のようなデータが一致することによって，写真を関連付く可能性がある．そこで，本研究では，デジタル資料館における写真と函館市史年表編，はこだて人物誌を LOD 化することで歴史的に関連する写真を関連付けることを目指す．

3. 歴史資料の LOD 化の検討

3.1 メタデータの調査

2.2 に記述したように，年月日や固有名詞などが資料間を関連付けるために重要なデータとなることがわかる．デジタル資料館における写真や函館市史年表編，はこだて人物誌を LOD として関連付けるためには，共通のデータを含んでいる必要がある．本稿では，デジタル資料館における写真と函館市史年表編，はこだて人物誌の年月日と人名の

表 1　写真のメタデータの例

タイトル	成聖式記念　函館ハリストス正教会　聖堂正門
内容説明	大正五年十月十五日挙行　成聖式記念絵葉書　函館ハリストス正教会（3 枚中 1 枚)、附録として解説「函館正教会由来」(1 枚）あり
印刷発行	函館ハリストス正教会
西暦	1916 年
和暦	大正 5 年
大きさ	
地域分類	北海道（函館）
資料番号	pc000766-001

データについて調査し，抽出方法について検討した．

3.2 写真のメタデータの調査と抽出方法

写真のメタデータは，表 1 にあるような 8 項目で構成されている．表 1 の大きさにあるように整備されていないメタデータが含まれている．また年月日や人名のみを記述している項目は持たず，タイトルと内容説明の中にそれらが含まれている．また，写真によって表記ゆれや年月日の意味が異なるものも含まれている．

年月日の表記の種類について次に記述する．括弧内は例を示す.

- 和暦で漢数字での表記 (昭和九年三月二十一日)
- 和暦で数字での表記 (昭和 30 年 5 月 10 日)
- 和暦で数字とピリオドでの表記 (10.7.1)
- 西暦で数字とカンマで日,月,年順の表記 (22,2,1912)

和暦で数字とピリオドで表記されている場合の年号は，和暦の項目に書かれている年号を確認する必要がある.また，年月のみ表記されているものも含まれている．

年月日の意味について次に示す.

- 写真に関連する出来事が起きた日付
- 写真が印刷・発行された日付
- 写真が検閲された日付

一つの項目の中に，写真に関連する出来事が起きた日付が複数含まれる場合もある．

以上から，デジタル資料館における写真の年月日をメタデータとして抽出するためには，テキスト処理で年月日または年月を抽出し，表記を統一し，年月日の意味で分類する必要がある．

3.3 年表のメタデータの調査と抽出方法

年表の年代は,図 1 にあるようにページ上部に西暦(和暦)で記述されている．また，それ以下に月日がピリオド区切りで記述され，月日に対応するように月日の右に出来事が記述されている．また，図 1 の“2.16”のように出来事の文章の中に日にちが含まれている場合がある．さらに“3.–”，“–.–”のように月のみ，月日のどちらも記述されていない場合も含まれている．

以上から，函館市史のデータを抽出するためには PDF をテキスト化し，テキスト処理で，西暦，和暦，年月日，出来事，出来事の中に含まれる日付を抽出する必要がある．

3.4 はこだて人物誌のメタデータの調査と抽出方法

はこだて人物誌とは，財団法人函館市文化・スポーツ振興財団の情報誌「ステップアップ」 に掲載された「函館ゆかりの人物伝」(1992.4 vol.37～2012 年現在も連載中）をもとに作成されている Web サイトである[5]．現在，約 260 名の函館にゆかりのある人物が紹介されており，約一ヶ月に一度，人物の追加更新がされている．Web サイトには，画像，人名，人名かな，生年，没年，概要文，詳細文，参考文献が含まれていた．これらのデータはスクレイピングにより抽出する必要がある．また，詳細文において人物に関連する出来事の年月日が多く含まれているためこれらの抽出方法については今後検討を進める予定である．

3.5 写真と年表の人物名の抽出

写真のタイトルや内容説明，年表にも人名は含まれている．そこで，はこだて人物誌の人名でそれらの文字列を検索しマッチした場合に抽出する．しかし，人名については資料間で表記ゆれしている場合があるため，共通の人物の人名をマッチさせるためには編集距離を適用するなど自然言語処理を用いることが必要であると考えられる．また，人物誌に含まれていない人名の抽出方法についても今後検討が必要である．

4. 市民参加型地域写真アーカイブの構築

市民の知識を利用した地域写真の関連付け手法の提案について述べる．

市民が情報提供の担い手となる地域写真アーカイブを構築するためには，市民が協調し，潜在している情報を引き出すための適切な場とインタフェースを設けることが必要である[6]．本研究では，SNS のような場を設け，市民が協調して写真についての知識やエピソードをアノテーションすることによって，被写体の情報を収集する．そしてそれらの情報を利用して関係の深い被写体間をシステムによって関連付ける仕組みの構築を目指す．特に市民によるアノテーションを利用した被写体の関連付け手法を提案する．

5. 被写体の関連付け手法

5.1 建造物の関連付け

本研究では，被写体として，地域写真に写されている建造物や石碑を対象とし関連付け手法の構築に取り組んだ．市民は建造物についてのエピソードや位置情報のような知識を持っている．また共通の建造物が含まれている二枚の写真が提示された場合に，それらが共通のものだと判断できると想定し，写真に含まれる建造物間の関連度を算出する手法を提案する．

5.2 エピソードの利用

市民が建造物に対して建造物に関するエピソードをアノテーションする．そのアノテーションされたエピソードを用いて，被写体間にアノテーションされているエピソードの関連度を算出する．

エピソードの関連度を算出する計算方法について説明する．エピソードに出現頻度の低い単語が共起する被写体は関連が強いとみなす．したがって，建造物のエピソードの関連度は以下の手順で計算する．

$$idf_t = 1 + \ln\left(\frac{\text{全てのエピソード数}}{\text{単語tが出現する回数}}\right) \tag{1}$$

式(1)は被写体のエピソード中に登場する単語の*idf*値の評価である．idf_tは，ある単語tが出現するエピソード数の逆頻度を示している．つまり，その共起した単語が出現するエピソード数が少ないほど，高い値をとる．

求める二つのエピソード間の*idf*値の総和*idfall*は次のように計算する．式(2)に計算式を示す．

$$idf_{all} = \sum_{k=1}^{n} idf_k \tag{2}$$

idf_kは，両エピソードで共起している単語の*idf*値である．これを，最も総和の値が高いエピソード間の*idf*値総和 $(idfall)max$ で除して正規化し，求める二つのエピソードの関連度r_eとする．式(3)に計算式を示す．

$$r_e = \frac{idf_{all}}{\left(idf_{all}\right)_{\max}} \tag{3}$$

総和を計算することで，共起した単語を含むエピソードが類似しているかを判断することができる．エピソードの類似性が高い場合，総和は高い値を示す．

5.3 位置情報の利用

市民が建造物に対して，緯度経度をアノテーションするアノテーションされた緯度経度を用いて，被写体間の距離から関連度を算出する．

距離による関連度を算出する計算方法について説明する．建造物間の距離を r とする．この距離 r を利用し，距離による関連度 r_g を計算する．式(4)に計算式を示す．

$$r_g = \begin{cases} 1 - \dfrac{r}{R} & (0 \le r \le R) \\ 0 & (otherwise) \end{cases} \tag{4}$$

式中の R は，距離が近いとみなす閾値を表す．R は，一つの被写体から，別の被写体を視認することができる距離で設定する．

5.4 直接関連付け

市民が複数の写真から関係があると判断できる建造物を見つけた場合にそれらの建造物を直接関連付ける．この関連度付けによる関連度 r_u を計算する．式(5)に計算式を示す．

$$r_u = \begin{cases} 1 & (\text{関連付けられた場合}) \\ 0 & (otherwise) \end{cases} \tag{5}$$

5.5 関連度の算出

エピソードによる関連度r_e，距離による関連度r_g，ユーザの直接指定による関連度r_uにそれぞれを重み付ける係数 w_e，w_g，w_u を掛け合わせることで被写体の関連度r_{sum}を計算する．式(6)に計算式を示す．

$$r_{sum} = w_e r_e + w_g r_g + w_u r_u \tag{6}$$

6. 評価実験

6.1 実験の目的と方法

3.5 で示した式(6)の重み付けを変化させることで市民の判断する写真の関連度に近似させることができるかを調べるため実験を行った．建造物を含む写真と建造物のエピソード，位置情報を利用し，w_e=w_g=w_u=1 の場合の写真の関連度を算出した．同時に，関係があると思われる写真のペアを被験者に可能な限り作成させた．実験には，函館市中央図書館が公開しているデジタルアーカイブの建造物を写した写真 10 枚を利用した(図 3). 写真には 1 から 10 の番号を付けた. 建造物のエピソードは, 写真のメタデータに「内容説明」の項目があるためそれを利用する．被験者は，函館在住期間約 7 年が 1 名，約 4 年が 3 名，約 5 年が 2 名の合計 6 名である．被験者には写真のメタデータも確認でき

図 3．評価実験用の写真

るようにした．被験者には関係していると判断した理由についても報告させた．

6.2 結果と考察

多くの被験者は共通の建造物を含む写真の間に関係があると判断した．また同じカテゴリに分類される建造物であることや，建造物の造りが似ていることで関連があると判断していた．結果，本手法によって被験者の判断に合う結果となったのは共通の建造物を含む写真のペアのみであった．市民は視覚的な情報からも関連性を判断することがわかった．

7. おわりに

本稿では，地域史を編纂するために，地域写真を歴史的関連で関連付ける手法について報告した．

手法の一つ目として，デジタル資料館における写真資料と函館市史年表編，はこだて人物誌の LOD による関連付けについて報告した．また，年月日と人名が各々を関連付けるために必要なデータであるとし，写真，年表，人物誌においてデータの調査をし，その抽出方法について検討した．今後，年月日，人名以外のデータでの関連付けについても調査を進める．そして外部の LOD ともリンクする RDF を作成するためにデータの表記の統一とデータ間を意味付けする語彙の選定を行う．さらに，関連付けられたデータの関連性についての評価も行う予定である．

手法の二つ目として，市民の知識を利用した地域写真の被写体の関連付け手法とその評価結果について報告した．関連付け手法は，市民のエピソード，位置情報，直接関連付けの三つの情報を用いる．そして各々の情報が複数の写真被写体にアノテーションされた場合に被写体間の関連度を算出する式を提案した．評価実験からは，地域についての知識が少ない市民であっても被写体の視覚的な情報から関連があると判断することがわかった．

今後は，LOD によって関連付けられた写真資料に，市民が私的写真や知識を付加することで歴史的関連性のある地域写真を関連付けていく仕組みについての検討を進める．

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ィジタルアーカイブ CMS，第 84 回デジタルドキュメント研究会 (2011).
2)　川嶋稔夫，木村健一：街の記録を編みあげるディジタルアーカイブ，第 23 回人工知能学会全国大会，pp.1 (2009).

3) 宮武志保，木村健一：地域コミュニティのストーリーテリングを支援する年表型巻物の提案，第 84 回デジタルドキュメント研究会 (2011).
4) 函館市図書館所蔵デジタルアーカイブ デジタル資料館, http://www.lib-hkd.jp/digital/
5) はこだて人物誌, http://www.city.hakodate.hokkaido.jp/soumu/hensan/jimbutsu_ver1.0/index.htm
6) 川嶋稔夫，木村健一：市民と編みあげる地域ディジタルアーカイブ，第 25 回人工知能学会全国大会，pp.4 (2011).