

統合利用を目的とした地域史資料の LOD 化

奥野 拓^{1,a)} 高橋 正輝² 山田 亜美¹ 川嶋 稔夫¹

概要: 本研究は様々な形式でデジタル化され散在する地域の歴史資料を LOD 化し、関連付けることにより統合利用を可能とすることを目的とする。本報告では、函館地域の例として、絵葉書の画像アーカイブ、地域史年表、歴史上の人物紹介をそれぞれ RDF データセット化した事例を取り上げる。また、それらを関連付けた統合利用例として、「エピソードでつながる函館歴史写真」と「函館ゆかりの人物スポット」を示す。

Making LOD of Regional Historical Records for Integrated Application

OKUNO TAKU^{1,a)} TAKAHASHI MASAKI² YAMADA AMI¹ KAWASHIMA TOSHIO¹

Abstract: Regional of historical records are digitized and on public via the Internet. This study aims at enabling integrated application of them by making LOD of them and link with each other. This report focuses on an activity in Hakodate, making RDF datasets of the archives of picture postcards, the chronology of city history, and the biography of famous historical persons. As examples of integrated applications, “The Historical Photographs of Hakodate Linked by Episodes” and “The Spots Associated with Persons of Hakodate” are shown.

1. はじめに

近年、地域の古文書や古写真、古地図などの歴史資料をデジタル画像として保存・蓄積するデジタルアーカイブ化が広く進められている。例えば、函館地域は全国的に見ても貴重な歴史資料が数多く残っており、保存と活用のためのデジタルアーカイブ化に積極的に取り組んでいる。公立はこだて未来大学のデジタルアーカイブ研究センターが函館市中央図書館と共同で継続的に開発・運用している函館市図書館所蔵デジタルアーカイブ「デジタル資料館」[1]がその筆頭であるが、その他にも、『函館市史』デジタル版[2]や、「はこだて人物誌」[3]などを函館市の Web サイトで公開している。このように、各地で様々な種類の地域史コンテンツが公開されている。これらは非常に貴重な地域史データであり学術的価値も高い。しかしながら、それぞれが無関係に独立して公開されている状態であり、史実として互いに補完し合う情報を含んでいるにも関わらず、十分

に活用できる状態になっていないのが現状である。

一方、ボトムアップ的アプローチにより Web 技術を用いた意味的関連性に基づくデータ関係を推進する取組みとして、Linked Open Data (LOD) が注目されている。国内でも鯖江市や横浜市などが積極的に行政や文化芸術など公共データの LOD 化に取り組んでいるほか、政府の IT 戦略本部は「電子行政オープンデータ戦略に関する提言」において、行政の保有する様々な分野のデータを 2 次利用可能な形式により公開し、民間のサービス等で活用可能とすることを求めている [4]。この提言では、地域に根ざした草の根的なオープンデータの展開が期待されており、その要となるのが地方自治体の保有するデータの積極的な公開である。その意味では、歴史資料を含む文化芸術資料も例外ではなく、市民による活用や他地域との関係など、LOD 化のメリットは大きいと考えられる。

本研究では、デジタルアーカイブに含まれる画像化された歴史資料、テキスト形式で公開されている地域史や年表、人物史などをデータ化し、相互に関連付けて組み合わせ、統合利用することにより新たな価値を創出することを目指す。LOD 化によるアプローチでは、相互に意味的関連付けされた RDF 形式のデータモデルを作成（データモデリン

¹ 公立はこだて未来大学
116-2, Kamedanakano-cho, Hakodate 041-8655, Japan

² 公立はこだて未来大学大学院
116-2, Kamedanakano-cho, Hakodate 041-8655, Japan

a) okuno@fun.ac.jp

グ) し、LOD として公開可能な地域史 RDF データセットを作成する。種類の異なる複数の地域史 RDF データセットを Wikipedia の LOD 版である DBpedia[5] などの外部リソースとリンクすることで、地域史データを多面的に活用できるようにすることが目標である。

このアプローチに基づき、本研究では函館地域を事例として、デジタル資料館に含まれる画像化された歴史資料、テキスト形式の歴史資料である函館市史年表編、及び、はこだて人物誌を対象とし、RDF データセットを作成している。また、それらを用いた 2 種類の統合利用例について検討を進めている。一つは、デジタル資料館の画像資料と函館市史年表編を組み合わせた「エピソードでつながる函館歴史写真」である。もう一つは、はこだて人物誌を函館市の公式観光情報サイトの観光情報コンテンツと組み合わせた「函館ゆかりの人物スポット」である。本報告では、以上の取り組みについて述べる。

2. 関連研究

文化芸術資料を対象とした LOD に関する研究としては、例えば以下のものがある。

萩原らは、目的に応じた適切なデジタルアーカイブの選択を支援することを目的とし、デジタルアーカイブ自体にメタデータを付与し LOD 化を行っている [6]。複数のデジタルアーカイブを対象としてデータ検索を行うという点では本研究テーマと共通する。

また、博物館・美術館情報などを LOD 化する試みとして、LODAC Project がある [7]。これは、様々な学術情報を関連付けて公開・活用することを目的としたプロジェクトである。本研究では、LODAC における語彙定義やデータモデリング手法を参考に、LODAC とのデータ関係を視野に入れたデータモデリングを行う。

これらの取り組みでは、デジタルアーカイブや所蔵品、作者、施設などの目録データの LOD 化が中心であるが、本研究では、函館市史などのテキスト形式の歴史資料の内容を対象とした意味的関連付けを試みる。そして、単純な全文検索では発見できないデータ間の関連性に基づくデータ検索を実現することを目指す。

3. 函館地域の歴史資料の RDF データ化

3.1 函館市図書館所蔵デジタルアーカイブ「デジタル資料館」

公立はこだて未来大学では函館市中央図書館と連携し、歴史資料のデジタル画像化を進めている。函館市中央図書館が所蔵する函館地域に関する歴史資料は、古写真や絵葉書、古文書、古地図、ポスターといった種別の資料が存在し、その総数は約 5 万点に及ぶ。デジタル画像化した資料の効率的かつ安定的な管理と公開を目的として、公立はこだて未来大学では、デジタルアーカイブ CMS を開発し、

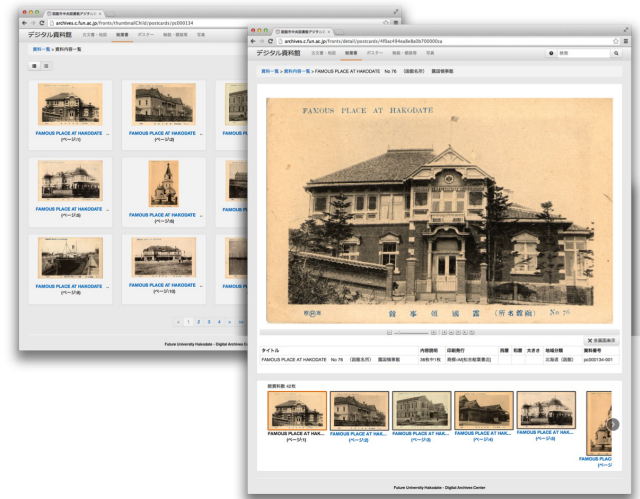


図 1 函館市図書館所蔵デジタルアーカイブ「デジタル資料館」
Fig. 1 Digital Archives of Hakodate City Library.

継続的にその改良を進めている [8]。デジタルアーカイブ CMS を用いて約 1 万 3 千点の高解像度画像と付随する目録情報を閲覧できる「デジタル資料館」が一般公開されている。図 1 に、デジタル資料館の資料一覧画面 (左) と資料詳細画面 (右) を示す。

画像資料を探す方法として、一覧画面よりサムネイル画像をブラウズするか、目録テキストの全文検索を行うことが可能であるが、これらの方法は膨大な点数のデジタルアーカイブから資料を見つけ出す方法として不十分である。資料間が共通の被写体や歴史的エピソードなどにより関連付けられれば、歴史的に関連する資料を容易に抽出することが可能となり、デジタルアーカイブの活用促進につながると考えられる。デジタル資料館の資料の目録に含まれる項目として、タイトル、内容説明、印刷発行、西暦、和暦、大きさ、地域分類、資料番号がある。これらは、デジタルアーカイブ CMS のデータベースにおいて項目ごとに管理されている。

目録のタイトルと内容説明の項目には、空白で区切られた複数の内容が記載されている。例えば、ある絵葉書はタイトルが「成聖式記念 函館ハリストス正教会 聖堂正門」であり、内容説明は「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」となっている。これらの内容には、日付や人名など、独立したメタデータとして扱うべき項目が含まれている。デジタル資料館の資料を他の様々な歴史資料と意味的に関連付けるためには、これらを適切に抽出する必要がある。

例えば、日付については、西暦と和暦の項目には年のみ記載されているのに対し、内容説明には年月日が記載されている。他の歴史資料との関連付けの際には、この年月日が重要な役割を果たすと考えられる。内容説明に含まれ

1912(明治45・大正1)
5905トンとなる。これを明治43年に比較すれば旅客は5万2312人、貨物は2万7928トンの増加となる[函日・明45, 1, 14]
2. 3 函館訓育院の創立発起人総会が商業学校内において開催され、七飯村に設置されることが決まる。大正13年4月、道庁に移管して序立大沼学院となる[函日・明45, 2, 1, 4-5, 新撰北海道史7]
2. 7 ドイツより帰朝した医師齊藤一郎の歓迎会が蓬萊町小林亭で開かれ、119人が出席する。席上、北守区長は満5年のドイツ滞在中に研鑽を費んだ細菌学と衛生学の研究が将来必ず函館区の公共に貢献するとの挨拶を述べる[函日・明45, 2, 9]
2. 10 函館八幡宮改築に関する八幡宮祭典委員・招魂社祭典委員・区会議員・有志らの相談会が市役所で行われる。経費5万円で大殿・中殿・拝殿を改築することが決まる[函日・明45, 2, 11]
2. 12 清国の宣統帝が退位し、袁世凱に臨時共和政府組織の全権を付与し、清朝は滅亡する。15日、南京参議院は袁世凱を中華民國臨時大總統に推薦し、3月10日に北京で就任式をあげる[近代日本通年表]
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ゲイ大主教が祭主を務める[函日・明45, 2, 20, 24]
2. 18 一昨日来の大吹雪で函館本線の二股・蔵宿間で列車が立ち往生し、黒松内・長万部からの救援機関車7台も進退きまわり約70人の乗客が列車内に閉じこめられる[函日・明45, 2, 19]

図2 函館市史年表編

Fig. 2 The chronology volume of the Hakodate City History.

る年月日をメタデータ化するためには以下の二つの課題がある。一つ目は、表記揺らぎである。和暦・西暦、漢数字・算用数字、区切り文字の違いの組み合わせが存在する。二つ目は、日付の意味付けである。例えば、ある出来事に関する写真の絵葉書の場合には、日付の解釈として、写真に関連する出来事が起きた日付、絵葉書が印刷・発行された日付、絵葉書が検閲された日付の可能性がある。記載内容からこれらを適切に解釈する必要がある。

3.2 『函館市史』デジタル版

函館市史は、函館の歴史をまとめた書籍であり、通説編(1~4巻)、資料編(1, 2巻)、年表編など、全7編11巻から構成される。このうち通説編の4巻については、『函館市史』デジタル版としてWeb公開されており、年表編についても現時点では非公開であるがPDF化されている。

年表編は、旧石器時代後期から2004年(平成16年)までの歴史年表であり、主に明治以降は新聞資料を基礎資料として作成されている。内容は、函館に関連する出来事の短い記述(本文)に西暦・和暦と月日(特定できないものを除く)および出典が付加されたものであり、約700ページにわたる。図2に、1912年の記述の一部を示す。年表編は、個々の出来事が簡潔にまとめられており、日付と対応付けられているため、他の歴史資料や観光スポットデータとの関連付けることで有効に活用できると考えられる。また、通説編がトピック単位の章立てとなっており、図表を含めた不定形の記述であるのに対し、年表編は表記法が統一されており、メタデータの抽出が比較的容易である。以上の理由により、函館市史のRDFデータ化については、まず年表編から着手している。年表編のPDF文書を作成したアプリケーションの文書ファイルが失われているため、PDF文書からのテキスト抽出から開始し、プレーンテキスト化された文書から凡例にしたがって日付、本文、出典を抽出し、データベースへの格納を行った。

3.3 はこだて人物誌

はこだて人物誌(以下、人物誌)は、函館市文化・スポー



図3 はこだて人物誌

Fig. 3 Hakodate Jimbutsushi (Famous historical persons of Hakodate).

ツ財団が毎月発行している広報誌「ステップアップ」[9]に連載中の「函館ゆかりの人物伝」を一部加筆修正してWebコンテンツ化し、函館市が公開しているものである。毎月1名ずつ追加しており、2013年9月1日時点では270名の函館に縁のある人物の歴史を記した貴重なコンテンツである。図3は「今井市右衛門」のページの一部である。各人物について、人名、かな人名、生没年、リード文、本文、人物の画像が掲載されており、本文にはその人物が誕生してから亡くなるまで、年代ごとに体験した出来事やその地名が記述されている。人物誌のコンテンツはHTMLにより構造化されて記述されているため、スクレイピング処理によりデータを抽出し、データベースへの格納を行った。

3.4 写真資料と年表の関連付け

3.1で述べた歴史的エピソードによる写真資料の関連付けを、函館市史年表編を用いて行うことを検討した結果、写真の目録と年表の出来事の本文に含まれる日付、人物、場所を用いた関連付けが有効であることがわかった[10]。そこで、写真の目録と年表の出来事の本文から、日付、人物、場所の情報の抽出を行った。

写真は昭和のものが大半であるため、年表は昭和(1926~1988)のものに限定した。写真における日付は、1で述べたように、記事内容から日付の意味を適切に解釈しなければならない。年表の日付とリンクするためには、写真の目録の中から出来事が起きた日付を抽出する必要がある。日付は、写真の目録において「タイトル」と「内容説明」に記述されていることが多い。その中でも出来事が起きた日付は、内容説明に記述されている傾向があることから、「内容説明」に含まれる日付をテキスト処理によって抽出した。

人物については、写真、年表ともに人物誌に掲載されている人物275名と文字列マッチすることによって抽出した。また、場所については、函館市公式観光情報サイト「はこ

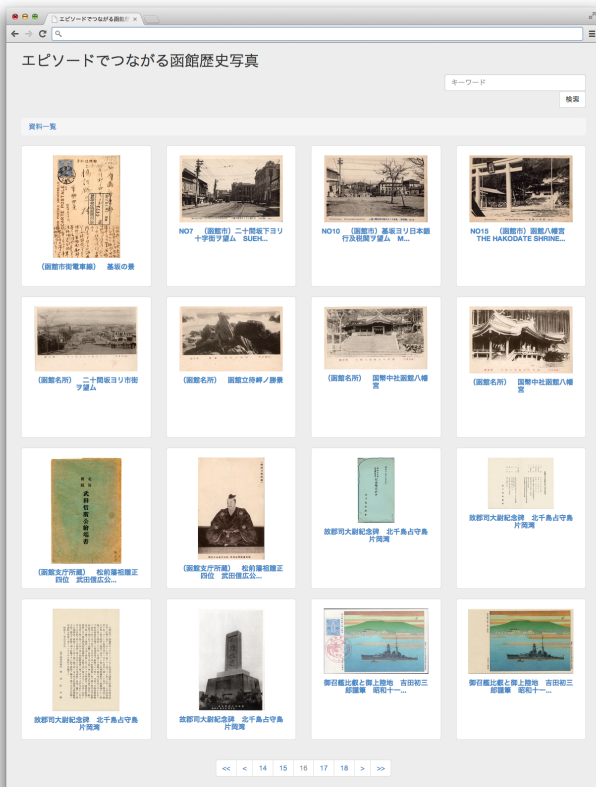


図 5 「エピソードでつながる函館歴史写真」の資料一覧画面
Fig. 5 A photo catalog view of “The Historical Photographs of Hakodate Linked by Episodes”.



図 7 「エピソードでつながる函館歴史写真」の資料詳細画面（関連画像一覧表示）
Fig. 7 A photo detail view of “The Historical Photographs of Hakodate Linked by Episodes” (with a catalog view of related photos).



図 6 「エピソードでつながる函館歴史写真」の資料詳細画面
Fig. 6 A photo detail view of “The Historical Photographs of Hakodate Linked by Episodes”.



図 8 「エピソードでつながる函館歴史写真」のエピソードでリンクされた画像表示画面
Fig. 8 An episode-linked photos view of “The Historical Photographs of Hakodate Linked by Episodes”.

させない」ことが念頭に置かれ、完成度の低い目録情報であっても資料と併せて公開している [15]。したがって、年

号や著者等の目録情報が完備されている資料がある一方で、目録が題名だけに留まる資料も存在する。

一方、地域の建造物やイベントなどを記録した写真資料については、市民が被写体に関する情報を持っている可能

性が高い [16]. そのため、観光資源への応用など、より広い目的での歴史資料の活用を考えた場合、市民参加型の編纂が有効であると考えられる。

そのひとつのアプローチとして、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し、関係の深い被写体を関連付ける仕組みについて検討している [17]. 特に、観光資源としての活用を目指し、現存する歴史的建造物や石碑などを対象とした関連付け手法について検討を行った [18].

しかし、現状、多くの市民は函館の歴史に興味を持つことがなく、写真が閲覧されること自体ほとんどない。そのような函館の歴史にあまり興味を持たない市民に対し、写真への興味を促すような仕組みとして、歴史写真閲覧システム「エピソードでつながる函館歴史写真」を構築した。これは、ユーザが写真間の歴史的関連性に基づき、興味のあるものを選択していくことで発見的に写真を辿って閲覧することができる Web アプリケーションである。現在、このシステムでは、3.6 で示したデジタル資料館の画像資料 RDF データセットと函館市史年表編 RDF データセットを使用している。

このシステムにアクセスすると、最初に図 5 のように、デジタル資料館と同様に、写真一覧が表示される。ここで、一覧ページをブラウズするか、あるいはキーワードを入力して全文検索することによって、興味のある写真を探す。1 枚の写真を選択 (タイトルをクリック) すると、写真の詳細情報が表示される。図 6 は、「(函館名所) 国幣中社 函館八幡宮」というタイトルの画像を選択して遷移した画面である。写真の下には、その写真の目録と、その写真に関連する年表の出来事 (エピソード) が年代順 (昇順) に表示される。ここで、橙色で表示されている文字列は、写真と年表を関連付けているリソースである。青色の文字列 (日付、人物、場所) を含む出来事は、それをリソースとして関連付けられた写真が存在していることを表す。その場合、出来事 (リンク) をクリックすると、図 7 のように、関連付けられた写真の一覧が出来事の直下に展開表示される。一覧の各写真の下の「選択した写真との関連」ボタンをクリックすると、図 8 のように、左側に最初に選択した写真、右側に関連写真、下にはそれらを関連付ける年表の出来事が表示される。

以上で示した一連の流れにより、最初に選択した函館八幡宮の絵葉書をきっかけにして「函館八幡宮には、昭和 11 年 10 月 10 日に陸軍特別大演習のために来道した昭和天皇が視察に訪れている」という史実と、それを記念して発行された絵葉書の存在を知ることができる。

4.2 函館ゆかりの人物スポット

本研究では、3.5 において示したように、人物にまつわ

る歴史的なエピソードを観光スポットと結びつけて旅行者に提示することにより、観光行動の喚起や満足度の向上を目指す。この目的においては、旅行中に訪れた観光スポットにおいて、リアルタイムに情報を提供できるような仕組みが有効であると考えられる。例えば、GPS 等により測位し、自己位置に基づいた情報提供を行うことが可能なスマートフォンアプリケーションである。アプリケーションの核となる RDF データセットの観光資源としての有効性を評価する目的で実験を行った。

コストをかけずに旅行者が観光行動中にアプリケーションを使用するのに近い状況を再現するために、人物誌 RDF データセットに含まれるデータと観光スポット情報を用いた Web サイト「函館ゆかりの人物スポット」を作成した (図 9)。

Web サイトは観光スポット毎のページ構成とし、ページの上部に観光スポットの情報を掲載し、その下に「このスポットにゆかりのある人物」として、その観光スポットに関連付けられた人物の紹介ブロックをタイル状に配置した。掲載した観光スポットは元町公園、大森浜の 2 箇所とした。観光スポットに関する情報としては、函館市公式観光情報サイト「はこぶら」の観光スポットの概要、説明、画像、住所、アクセス、カテゴリを掲載した。「このスポットに緑のある人物」には、人物誌に掲載されている人物の概要、画像、観光スポットと関連のある出来事、関連のあるスポットを掲載した。

この Web サイトを用いて以下の手順で実験を行った。まず被験者に各観光スポットを散策している様子が映っている 3 分間の動画を視聴させる。次に、観光客として訪れた状況を想定したシナリオを読ませることで、函館に訪れた観光客の視点に立つよう促し、被験者に Web ページの観光スポットと、関連のある人物の紹介を読むよう指示する。また、元町公園に関しては、元町公園内に設置されている「四天王像」がシナリオに関わっているため、四天王像とその解説板の画像を提示する。その後、人物情報によって観光スポットへの興味が深まったかどうかという観点のアンケートを行う。

実験は函館市内に住む 12 名、函館市外に住む 2 名の学生を被験者として行った。アンケートの結果、「人物情報から新たな発見があったか?」という質問に対し、「大きな発見があった」「発見があった」という回答が約 80%であった。自由記述では、「観光していても歴史について調べる機会がなかったため、このような人物が関わっていることは知らなかった」などの意見があった。「関連のある出来事によって観光スポットへの興味が深まったか?」という質問に対しては、「とても興味が深まった」「興味が深まった」という回答が約 70%という結果となった。この実験より、作成した人物誌 RDF データセットが観光資源として有効であることが伺える。



図 9 「函館ゆかりの人物スポット」の元町公園ページ

Fig. 9 The Motomachi-Koen page of “The Spots Associated with Persons of Hakodate”.

5. おわりに

本研究では、地域に散在するの様々な歴史資料を LOD という形で公開し、相互に連係させることで、観光を始めとする様々な分野で統合利用することを目指した取り組みを行ってきている。本報告では、函館地域の例として、絵葉書の画像アーカイブ、地域史年表、歴史上の人物紹介をそれぞれ RDF データセット化した事例を取り上げ、それらを関連付けた 2 種類の統合利用例を示した。

現時点では、比較的抽出が容易な、日付、人物名、場所（観光スポット）名をリソースとしたデータ間の関連付けを行っている。今後は、より効果的な統合利用に向けて、非構造化テキストから様々な意味的関連性を抽出すること

を検討する予定である。

謝辞 本研究の推進にあたって全面的に協力して頂いている函館市中央図書館、そして函館市公式観光情報サイト「はこぶら」の観光スポットデータの利用許可を頂いた函館市観光コンベンション部の皆様へ感謝致します。

参考文献

- [1] デジタル資料館, 函館市図書館所蔵デジタルアーカイブ (online), 入手先 <<http://archives.c.fun.ac.jp/>>
- [2] 『函館市史』 デジタル版 (online), 入手先 <http://www.city.hakodate.hokkaido.jp/soumu/hensan/hakodateshishi/shishi_index.htm>
- [3] はこだて人物誌 (online), 入手先 <http://www.city.hakodate.hokkaido.jp/soumu/hensan/jimbutsu_ver1.0/>
- [4] 電子行政オープンデータ戦略に関する提言 (online), 入手先 <<http://www.kantei.go.jp/jp/singi/it2/denshigyousei/honbun.pdf>> (2011)
- [5] DBpedia Japanese(online), 入手先 <<http://ja.dbpedia.org>>
- [6] 萩原和樹, 中野伸吾, 本間維, 永森光晴, 杉本重雄: Linking Digital Archives: LOD を利用したデジタルアーカイブへのアクセス支援, 情報処理学会第 75 回全国大会講演論文集, 4N-2, 1-587 (2013)
- [7] LODAC Project(online), 入手先 <<http://lod.ac>>
- [8] 出口貴也, 中原裕成, 高橋正輝, 奥野拓, 川嶋稔夫: 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, 情報処理学会研究報告, Vol. 2011-DD84 (2011)
- [9] ステップアップ (online), 入手先 <<http://www.zaidan-hakodate.com/stepup/>>
- [10] 高橋正輝, 奥野拓, 川嶋稔夫: 函館の歴史資料を用いた地域写真アーカイブの編纂, 情報処理学会研究報告, Vol.2013-DD88 No.9 (2013)
- [11] 函館市公式観光情報サイト「はこぶら」 (online), 入手先 <<http://www.hakobura.jp/>>
- [12] 堀井洋, 林正治, 堀井美里, 沢田史子, 吉田武稔: 歴史資料が有する観光的特徴の分析とその活用 「梅田日記」を事例として, 情報処理学会研究報告, Vol. 2009-CH83, pp. 1-7 (2009)
- [13] 山田亜美, 高橋正輝, 奥野拓: 函館地域史データの観光資源化に向けた地域史 LOD の構築—人物誌を対象とした文章からの歴史情報抽出と観光情報との関連付け—, 情報処理北海道シンポジウム 2013 講演論文集 (2013)
- [14] 石川徹也, 北内啓, 城塚音也: 歴史オントロジー構築のための史料からの人物情報抽出, 自然言語処理, Vol.15, No. 1, pp. 3-18 (2008)
- [15] 川嶋稔夫: 函館の写真アーカイブとコミュニケーション, 情報処理学会研究報告, Vol. 2008-DD69, No.7 (2008)
- [16] 川嶋稔夫, 木村健一: 市民と編みあげる地域デジタルアーカイブ, 第 25 回人工知能学会全国大会, 2A1-OS15-7 (2011)
- [17] 高橋正輝, 奥野拓: 地域写真アーカイブにおける共通する被写体の関連付け手法, 情報処理北海道シンポジウム 2011 講演論文集, pp. 103-104 (2011)
- [18] 高橋正輝, 奥野拓: 観光情報サービスでの活用に向けた地域写真アーカイブにおける建造物の関連付け手法, 観光情報学会第 5 回研究発表会講演論文集, pp. 1-5 (2012)