

令和 5 年 6 月 16 日現在

機関番号：20103

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K12547

研究課題名(和文) 地域デジタルアーカイブの横断利用による地域学習教材作成支援システムの構築

研究課題名(英文) Construction of a Support System for Creating Regional Learning Materials through Cross-Use of Regional Digital Archives

研究代表者

奥野 拓 (Okuno, Taku)

公立はこだて未来大学・システム情報科学部・教授

研究者番号：30360936

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究では、地域学習教材の作成支援を目的として、複数の地域デジタルアーカイブを横断曖昧検索することにより、地域学習のテーマに関連する資料を容易に網羅的に検索し、資料の主題となる場所を古地図上に自動表示する手法を構築した。また、地域文化財アーカイブを対象として、説明文の特徴語に基づく可視化により、全体像の把握を容易にし、探索を支援する手法、画像資料アーカイブを対象として、画像の類似性およびメタデータの類似性に基づいて、資料の探索を支援する手法、博物館において、展示されている資料から、デジタルアーカイブ中の関連する所蔵資料の探索を支援する手法をそれぞれ構築した。

研究成果の学術的意義や社会的意義

デジタルアーカイブの資料の探索には、キーワード検索、階層的カテゴリ分類、タグ付けなどの方法が用いられるが、ユーザの知識を前提としたり、人手によるメンテナンスコストが高いという問題がある。本研究では、自然言語処理技術や深層学習技術を用いて、単語の意味的類似性を考慮した曖昧検索や、テキストや画像など資料の種類に応じた可視化手法を構築することにより、それらの問題の解決を試みているという点で学術的・社会的意義がある。

研究成果の概要(英文)：In this study, for the purpose of supporting the creation of regional learning materials, we constructed a method to easily and comprehensively search for materials related to the theme of regional learning by cross-ambiguous search across multiple regional digital archives, and to automatically display the location of the subject of the materials on an old map. In addition, we have developed a method to support search of regional cultural heritage archives by visualization based on characteristic words of explanatory text to facilitate understanding of an overall picture of the archived items, a method to support search of image archives based on similarity of images and metadata, and a method to support the search for related materials in the digital archive from the exhibited materials in a museum.

研究分野：デジタルアーカイブ、観光情報学、ソフトウェア工学

キーワード：地域学習 デジタルアーカイブ 探索支援 可視化 自然言語処理 深層学習 Webアプリケーション
スマートフォン

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1 . 研究開始当初の背景

小学校 3・4 年生の社会科では、副読本を用いて郷土の文化、伝統、生活などを学習することが一般的である。また、総合学習においても、地域社会に存在する多様な学習資源を直接的に体験することが学びに繋がる。児童が地域の歴史に興味を持ち、主体的に学んでいくことを促すためには、調べ学習を通して自ら発見し、歴史資料の閲覧や現地学習などを通して体験するというプロセスが重要である。これらの学習場面において活用可能な地域学習教材として、地域デジタルアーカイブは大きなポテンシャルを持つ。地域デジタルアーカイブを活用して授業を構成するためには、教員が様々な地域史コンテンツから横断的に関連する資料を探し、予め教材化しておく必要がある。しかしながら、関連する資料が様々なアーカイブに分散していて、そのままでは教材として使い難い。教師が集約して教材化する負荷は大きく、生徒の興味や関心に基づく発展的な学習に対応することは難しい。教師が効率的に効果的な教材を作成するためには、複数の地域デジタルアーカイブからテーマに沿った題材を探索する作業を支援する必要がある。

2 . 研究の目的

本研究では、複数の地域デジタルアーカイブの横断利用により地域学習教材を作成する作業を支援する枠組みを構築することを目的とする。アーカイブ内のコンテンツをキーワードにより関連付け、相互にジャンプできるようなシステムは既に実運用されている。しかし、このような文字列一致による関連付けだけでは、たとえば、「教会」と「修道院」のように意味的に関連するようなキーワードによってコンテンツを結びつけることができないため、意味的に関連するトピックを網羅的に探すことは困難である。一方、共通するキーワードが含まれていても、トピックとしては無関係な場合もある。独立して公開されている複数の地域アーカイブ間については、そのような単純なキーワードによる関連付けであっても管理主体が異なるため事実上不可能である。そこで本研究では、コンテンツに含まれる単語などの間の意味的な類似性に基づいて複数のアーカイブのコンテンツを動的に関連付ける手法を構築する。そして、独立した状態で公開されている自治体史や文化財情報、古地図・古写真、人物史などの地域史コンテンツを集約し、関連性に基づいて閲覧できるようにすることにより、地域学習教材の作成を支援するシステムを構築する。一方、テーマの手掛かりとなるようなキーワードなどが無い状態で、地域デジタルアーカイブをブラウジングしながら題材を探すようなケースでは、地域デジタルアーカイブの資料全体を俯瞰し、そこから興味を持つ資料を発見するような探索的な過程となる。そのような過程を支援するためには、資料全体を可視化することが重要であるが、可視化の手法は、テキスト主体のアーカイブか、画像主体のアーカイブかなど、アーカイブされている資料の種類に依存する。そこで、本研究では、アーカイブの種類や、探索の場面に応じた可視化による探索手法を構築する。

3 . 研究の方法

本研究は以下の 4 つのサブテーマに分けて実施した。(1)は、複数の地域デジタルアーカイブからの横断検索手法および古地図による可視化手法の構築、(2)～(4)は、資料の種類に応じた可視化による探索手法の構築である。

(1) 複数の地域デジタルアーカイブを横断曖昧検索することにより、地域学習のテーマに関連する資料を容易に網羅的に検索する手法、および、資料の主題に関連する地名を自動推定して古地図上に場所を表示する手法を構築する。

(2) 地域文化財アーカイブにおいて、アーカイブに含まれる様々な種類の文化財の特徴語に基づく可視化により、全体像の把握を容易にし、文化財の探索を支援する手法を構築する。

(3) 画像資料アーカイブにおいて、画像の類似性およびメタデータの類似性に基づいて、資料を探索することを支援する手法を構築する。

(4) 博物館において、展示されている資料から、デジタルアーカイブ中の関連する所蔵資料を探索することを支援する手法を構築する。

4 . 研究成果

(1) 本研究では、実際に教員に使用してもらうことを想定し、Web アプリケーションとしてシステムを実装している。現時点では二つの画面を実装している。一つ目の画面は、歴史的な人物や歴史的な出来事などの検索ワードにより複数の地域デジタルアーカイブから関連する資料を検索する「横断曖昧検索画面」である(図 1)。二つ目の画面は、横断曖昧検索画面の検索結果から選択された資料の説明文を表示し、説明文に記述されている歴史的な出来事が起きた場所を古地図上に表示する「資料詳細・古地図画面」である(図 1)。検索対象とする地域デジタ

ルアーカイブは、「函館市史デジタル版」、道南地域にゆかりのある歴史上の人物を紹介する「函館ゆかりの人物伝」、道南地域の文化財をマップ紹介する「南北海道文化財」、函館市中央図書館が所蔵している古写真や古文書、絵葉書、ポスターなどの画像資料を公開する「函館市中央図書館デジタル資料館」である。

横断曖昧検索画面では、類義語や表記揺れに対応させるため、Word2Vec を用いて検索ワードと類似度の高い単語を求め、資料の説明文中にそれらが含まれていれば、検索結果に含めるという処理を行う。

資料詳細閲覧画面の左側には、オープンソースの古地図ビューアライブラリ Maplat を用いて古地図を表示する。歴史的な出来事が起きた場所を古地図上に表示するための前処理として、文化財名と町名の収集、ジオコーディング、幾何補正を行う。当時の街並みや地形から歴史的な出来事についての理解をより深めるために、歴史的な出来事が起きた年代と最も近い年代の古地図を自動選択して表示する。現時点で実装しているシステムでは、「函館市中央図書館デジタル資料館」に収録されている明治 11 年から昭和 27 年の間の 9 種類の函館の古地図を用いている。

資料詳細閲覧画面の右側上部には、資料の説明文と、関連資料を表示する。説明文に含まれる文化財名と町名をハイライト表示し、クリックすると、古地図上にその場所をマーカー表示する。多くの資料には文化財名と町名が複数記述されているため、歴史的な出来事が起きた場所を特定することが困難である。そのため、資料に記述されている文化財名と町名を抽出し、文脈からどこで歴史的な出来事が起きたかを推定する。地名の推定は、地名候補の抽出、位置情報の特定、関連語の抽出、主題となる地名の推定の順で行う。そして、推定した文化財名または町名を歴史的な出来事が起きた場所として地図上にマーカー表示する。

資料詳細閲覧画面の右側下部には、閲覧中の文化財に関連する文化財をリスト表示する。本研究では、資料同士を関連付けるために、資料の内容を特徴付けることができる特徴語を用いる。例えば、人物資料の「時任為基」と文化財資料の「大噴水塔」という二つの資料の文章中には、「上水道」という特徴語が存在する。このような特徴語を TF-IDF により自動的に抽出し、同じ特徴語が存在する資料同士の関連付けを行う。



① 地域デジタルアーカイブ横断曖昧検索画面

② 資料詳細・古地図画面

図 1 地域教材作成支援システム

(2) 道南地域の学芸員団体が運営する「南北海道文化財」にアーカイブされている文化財には、メタデータとしてカテゴリが付与されているが、粒度が大きく、関連する文化財を抽出する目的には不十分である。そこで本研究では、文化財のタイトルと説明文から特徴語を抽出し、意味的類似性に基づいて特徴語の階層的クラスタリングを行い、類似する文化財を抽出する。そして、特徴語クラスタの階層構造を視覚化することにより、関連する文化財を探索できるようにする。

特徴語クラスタは、(i) TF-IDF によるタイトルと説明文からの特徴語抽出、(ii) Word2Vec を用いてベクトル表現された特徴語の階層的クラスタリング、(iii) 日本語 WordNet および Wikipedia を用いて抽出された特徴語の上位概念によるクラスタのラベリング、(iv) タイトルと説明文から抽出した名詞とクラスタを形成する特徴語とのコサイン類似度の近さに基づくクラスタと文化財との関連付け、という手順で作成する。以上の手順により作成した特徴語クラスタを、階層構造のデータ視覚化手法の一つであり、データ視覚化ライブラリ D3.js に含まれている Zoomable Circle Packing により視覚化する。この手法では、任意の円を選択することによって、その階層にズームされ、対話的に階層を辿ることができる。初期状態で表示されるクラスタ階層の位置を調節することにより、全体のクラスタ分布を俯瞰しながら探索できる。クラスタに含まれる文化財の数を円の面積に対応させることにより、クラスタの規模も同時に把握できる。

例えば、ユーザーが「山津波犠牲者慰霊碑」のページ(図 2)にアクセスした場合、視覚化ペインには「山津波犠牲者慰霊碑」が関連付けられた特徴語クラスタである「墓、岬」、「事故、戦闘」などが表示される。ここで、ユーザーが視覚化ペインの左下に位置する「事故、戦闘」クラスタをクリックすると、「事故、戦闘に関する文化財」を関連度の降順にタイル表示するページ(図 2)に遷移する。「事故」についてより詳しく調べたい場合は、さらに下の階層の「事

故」クラスタをクリックすることによって、「事故に関する文化財」ページに遷移し、一覧表示される文化財が「事故」のみに関連するものに絞り込まれる。閲覧中のアイテムに関連するアイテムのリストを表示する推薦システムは多数存在するが、提案システムには、関連の方向性をユーザーが把握し、選択しながら探索できるという特徴がある。



図2 特徴語の視覚化による文化財探索支援システム

(3) デジタルアーカイブを利用する際、閲覧したい特定の歴史資料がなく、ブラウジングして興味のある資料を探索する場合がある。この場合、公開されている資料の全容を把握できていると、興味のある資料の発見が容易になる。しかし、膨大な資料が公開されていると、カテゴリやタグといった方法では大まかにしか絞り込めない場合がある。その場合、何回もページを遷移し、資料をブラウジングする必要があるし、興味のある資料を発見することは容易ではない。そこで本研究では、画像資料を対象とした地域デジタルアーカイブである函館市中央図書館「デジタル資料館」において公開されている資料の全容の把握を容易にし、興味のある資料の発見を支援するシステムを構築した。

本システムでは、畳み込みニューラルネットワーク Inception-v3 の識別層の手前にある第3プーリング層からの出力を画像の特徴ベクトルとして抽出し、資料の画像間の類似度を算出する。そして、自己組織化マップ (SOM) を用いて類似度の高い画像を持つ資料が近接するような配置を行う。これにより、似た特徴を持つ資料の把握を容易にする。しかし、膨大な資料を収蔵するアーカイブでは、一つの画面に全ての画像を表示することは現実的ではない、そこで、段階的に探索可能にするために、画像の類似性に基づく配置から代表資料群を選出する。そして、最初に代表資料群を提示し(図3 ①)、そこから類似資料(図3 ②)へと探索可能にする。代表資料の選出手順として、まず、前述した画像の類似性に基づく配置結果を3×3の格子に分割し、格子内に含まれる画像の特徴ベクトルの平均ベクトルを算出する。そして、格子内において平均ベクトルに最も近い特徴ベクトルの画像を持つ資料を、格子内の代表資料として選出する。

画像の類似性によって配置するだけでは、画像間の類似性はないが関連がある資料が遠方に配置されることがある。そこで、選択された資料を基準に、資料に付与されたメタデータの類似性に基づく配置に並び替え可能にすることで、内容的に関連する資料の把握も容易にする(図3 ③)。本研究では、メタデータのうち資料の内容を端的に表しているタイトルを用いる。

以上の方法により、画像特徴の類似性に基づく探索と、タイトルの類似性に基づく探索を行き来しながら、興味のある資料を探索することが可能となる。



図3 画像とタイトルの類似性に基づく歴史資料探索支援システム

(4) 博物館では、歴史的背景や情景、年代、分類など、ある決められた観点に則って資料が展示されている。そのため、類似性や関連性のある資料を容易に閲覧することが可能である。しかし、類似性や関連性があり、共に展示することが望ましいにも関わらず、展示スペースの制約から展示されていない資料も多く存在する。日本の博物館では収蔵資料の3/4割しか展示しておらず、展示したことのある資料の割合の平均値は41.3%、中央値は30.0%である。そのため、博物館のデジタルアーカイブを活用し、展示されていない資料を閲覧可能にし、閲覧中の資料と類似性や関連性のある資料の鑑賞を容易にするスマートフォンアプリケーションを開発した。スマートフォンのカメラ機能を利用して閲覧中の資料を特定し、展示されている資料と類似性、関連性のある資料を表示することで、来館者の資料探索を支援する。本研究では、市立函館博物館を対象に、市立函館博物館デジタルアーカイブを活用したシステムを開発した。

開発したアプリケーションの利用例を下図に示す。このアプリケーションでは、閲覧中の資料の近くに設置したQRコードを読み取ることで資料を特定する(図4)。資料を特定すると、画面上部に閲覧中の資料を表示し、その下に摘要と資料名から算出された類似度の高い順に市立函館博物館デジタルアーカイブ上の資料を表示する(図4)。閲覧中の資料をタップすることで、市立函館博物館デジタルアーカイブに登録されている閲覧中の資料の詳細ページを表示する(図4)。類似性のある資料を表示する画面は、資料間の類似度を可視化するために、サムネイル画像と共に類似度の値に応じたゲージを表示する。類似度の値に応じたゲージを見ることで、来館者は閲覧中の資料と一覧で表示された資料の間にどの程度の類似性があるのかを知ることができる。類似性のある資料をタップすると、市立函館博物館デジタルアーカイブに登録されている資料の詳細ページを表示する。上部のタブにより、類似度、関連度を算出する際に用いたメタデータごとに表示する資料を切り替えることができる。



図4 博物館におけるデジタルアーカイブを活用した鑑賞支援システム

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 奥野拓, 宮井和輝	4. 巻 17
2. 論文標題 地域文化財デジタルアーカイブを活用したコンテンツ自動生成による歴史観光アプリケーションの構築	5. 発行年 2021年
3. 雑誌名 観光と情報	6. 最初と最後の頁 53-68
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 奥野拓, 長内一真	4. 巻 17
2. 論文標題 スマートフォンとWi-Fiによる函館リトファスゾイレの閲覧・周遊支援	5. 発行年 2021年
3. 雑誌名 観光と情報	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 高橋奎太, 奥野拓
2. 発表標題 Webサイト上の地域史資料を活用した地域教材作成支援システムの構築：初等教育における地域学習を対象として
3. 学会等名 デジタルアーカイブ学会第1回DAフォーラム
4. 発表年 2021年

1. 発表者名 奥野拓
2. 発表標題 地域文化財のデジタルアーカイブ化とオープンデータ 化による活用の試み：「南北海道の文化財」の事例
3. 学会等名 デジタルアーカイブ学会第5回研究大会
4. 発表年 2020年

1. 発表者名 高橋奎太, 奥野拓
2. 発表標題 Webサイト上の地域史資料を活用した地域教材作成支援システムの構築
3. 学会等名 情報処理学会情報処理北海道シンポジウム 2020
4. 発表年 2020年

1. 発表者名 高橋奎太, 奥野拓
2. 発表標題 Webサイト上の地域史資料を活用した地域教材作成支援システムの構築
3. 学会等名 情報処理学会第83回全国大会
4. 発表年 2021年

1. 発表者名 田中駿平, 奥野拓
2. 発表標題 画像とタイトルの類似性に基づく歴史資料探索支援システムの構築
3. 学会等名 情報処理学会第85回全国大会
4. 発表年 2023年

1. 発表者名 浦田柁, 奥野拓
2. 発表標題 博物館におけるデジタルアーカイブを活用した鑑賞支援システムの構築
3. 学会等名 情報処理学会第131回人文科学とコンピュータ研究会発表会
4. 発表年 2023年

1. 発表者名 奥野拓
2. 発表標題 特徴語の視覚化による文化財探索支援システムの構築
3. 学会等名 デジタルアーカイブ学会第7回研究大会
4. 発表年 2022年

1. 発表者名 田中駿平, 奥野拓
2. 発表標題 画像およびタイトルの類似性に基づく歴史資料探索支援システムの構築
3. 学会等名 情報処理北海道シンポジウム2022
4. 発表年 2022年

1. 発表者名 浦田柁, 奥野拓
2. 発表標題 デジタルアーカイブを活用した博物館における観賞支援システムの開発
3. 学会等名 情報処理北海道シンポジウム2022
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	川嶋 稔夫 (Kawashima Toshio) (20152952)	公立はこだて未来大学・システム情報科学部・教授 (20103)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------