

令和 6 年 6 月 25 日現在

機関番号：32605

研究種目：基盤研究(C)（一般）

研究期間：2018～2023

課題番号：18K00972

研究課題名（和文）多様な古文書字形および花押デジタルアーカイブをユビキタスに結ぶ情報検索技術

研究課題名（英文）Ubiquitous information retrieval technologies for digital archives of historical characters and Kao signatures.

研究代表者

末代 誠仁（Kitadai, Akihito）

桜美林大学・リベラルアーツ学群・准教授

研究者番号：00401456

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究は、古文書字形、花押などの図形情報を対象とした情報検索技術の高度化・高次化を目指したものであった。この検索技術は、パターン認識技術の活用によってメタデータからの独立性を実現し、使用言語や文化的背景が異なる多くの人々がデジタルアーカイブを通して連携することを可能にするためのものであった。

しかし、研究期間中に発生したCOVID-19のパンデミックにより研究計画の変更を余儀なくされた。そこで、上記検索技術を応用し、AIリテラシー獲得のための教材および教示法の開発を目指した。大学生数十名の実験において、同教材および教示法がAI学修に対する学び手のモチベーション向上に資する可能性が示唆された。

研究成果の学術的意義や社会的意義

情報検索技術の高度化に関しては、いくつかの研究発表を通して、画像をキーとした古文書／古典籍デジタルアーカイビング検索のための情報技術の価値を示すことができた。メタデータの利活用を前提とした古文書／古典籍デジタルアーカイビングの流れに一石を投じることができたと考えている。また、教材・教示法の開発については、実証実験を通して効果の可能性を示すことができた。この成果は、文部科学省を中心に実施されている「数理・データサイエンス・AI教育プログラム」の趣旨にも沿うものであり、詳細な検証を通して今後の教育活動への展開が期待できるものと考えている。

研究成果の概要（英文）：The aim of this research plan was to improve the accuracy of information search methods for historical character patterns and Kao signatures. Since the search methods did not depend on metadata sets on digital archives of the historical properties, they would support communications and collaborations between users of the digital archives of the historical properties.

Unfortunately, because of the COVID-19 pandemic occurring in the period of the research, the aim was modified. The new aim was utilizing the technologies of the search methods to develop teaching materials and methods for learners to obtain AI literacies. The results of the preliminary experiment employing dozens of students suggested the efficiency to improve the learners' motivation to learn AI.

研究分野：情報科学

キーワード：情報検索 古文書 花押 情報教育 AIリテラシー

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

本研究課題の申請時は、古文書をはじめとする歴史的な遺物のデジタルアーカイビングをより実質的な水準に昇華させ、幅広く共有することの重要性が特に強く認識されるようになった時期であった。歴史的に見ると、現在の国境線や使用言語の境界とは異なる形で文化・言語が共有された時代の存在は大前提である。この点に照らし、歴史学の研究については当該遺物を含む当時の言語・文化圏を重視した協力体制の下で実施することが重視されるようになっていた。国内においては自治体単位で運用されている教育委員会と国立（既に独立行政法人化されているが設立時に託された役割を示すためこの表現を用いる）の各研究機関との連携、および、国際的な協力関係の下で研究をボーダーレス/ユニバーサルに実施していくことが重要であるとの認識が、広く強く認識された時期に、当研究課題は申請・開始された。

遺物のデジタルアーカイビングに関する研究では、遺物の姿を画像や3Dモデルとしてデジタル情報化することに加えて、メタデータの整備が大きな課題となっていた。特にメタデータの整備においては、現在の自治体や国の境界線を越えた知見の構築につながる発見が必要であった。その発見につながる現在の研究者同士の交流が生み出す様々な情報を、パターン認識の技術の中核にした情報検索技術と有機的に結合し、メタデータの整備・拡充に資する情報技術の構築・提供を目指したのが本研究課題であった。

パターン認識技術を用いた情報検索は、キーワードやメタデータのように人間が抽象化した情報から独立した動作が可能である。一般的に、検索エンジンに入力する検索キーワードやデジタルアーカイブ上のメタデータは利用者/提供者となる人間が抽象化した情報だが、この抽象化の手続き/結果は抽象化を担った人間の知識、価値観、利用言語といったバックグラウンドに影響を受ける。前述したボーダーレス/ユニバーサルな研究を支援するにあたっては、国/地域ごとの利用言語（方言を含む）や注目する時代などの違いを吸収/超越する方法が求められるが、抽象化の手続き/結果がこの点を満足することは容易ではない。検索オプションの拡充により抽象化の粒度や方向性に变化を与え、検索キーに立体的な情報表現力を与えることで上記の問題を解決しようとする試みは様々な情報検索システムで実践されているが、抽象化された情報を利用する以上、本質的な改善とはいえない。一方で、パターン認識では、例えば「画像という事象そのものをキーとして他の画像を検索する」といった、利用者による抽象化を介さない情報検索が可能になる。

もちろん、字形/花押などが記された古文書は、それらの筆者が何らかの事象を抽象化した情報と捉えることができる。このような筆者自身による抽象化を事象と結び付けるのは、我が国においては考古学の領域に入る課題と考えられるが、本研究課題はそのような当時の抽象化の手続きを直接扱うものではない。本研究課題では遺物を事象そのものとして捉えながら、筆者以外の人間が実施する抽象化を可能な限り取り除いた場合の情報検索技術を実装・検証することで、遺物に忠実でボーダーレス/ユニバーサルな歴史学の実現に資する成果の達成を目指したものであった。

なお、申請者は過去の科学研究費助成事業を含む様々な研究活動の中で、本研究課題の土台となる文字画像検索の基礎的な技術を確立していた。本申請課題では、その高度化と拡充が具体的な研究の柱となる計画であった。

2. 研究の目的

(1) 期間開始当初の研究目的

本研究課題の初年度となる2018年度における研究活動の具体的な目標は、歴史学の研究者がボーダーレス/ユニバーサルな協力を通して得た様々な情報をパターン認識技術に織り込み、歴史学に関する幅広い活動に利用してもらえる文字/文字列画像検索技術の構築であった。ここには機械学習手法の利活用も含まれていた。申請者自身も、東京農工大学中川正樹研究室との協力の下で目的の達成を目指した。また、奈良文化財研究所、東京大学史料編纂所などの歴史学研究者の協力を得て、目標となる検索技術の性能向上を図ることも重要な目的であった。

(2) 社会の変化と研究目的の修正

しかし、2019年末にCOVID-19の世界的な感染拡大が発生し、本研究課題の基盤となる歴史学研究者の交流に変化が生じることになった。その一方で、TEIに代表される歴史学情報のマークアップと人工知能（以下、AI）に関する技術の連携が歴史学領域において広く論じられるようになり、本研究課題が当初目指した抽象化の手続きを介さない情報検索技術の研究についてはいったん棚上げを余儀なくされた。

その一方で、AIの利活用が人類にとっての社会的な責任と化し、AI技術によるシンギュラリティの到来が人材育成、教育などの分野に大きな影響を及ぼすようになる中で、申請者が本研究課題および他の研究活動で実現してきた情報技術を教育現場で活かすという新たなアイデアとニーズが発生した。そこで、本研究課題では研究期間の延長に関する科学研究費助成事業のルールも慎重に活用しつつ、古文書字形や花押を学習/学修対象のコンテンツとして活用でき、かつ、学問領域の志向性を問わない幅広い学習者/学修者がAIリテラシーを身に着けることのできる

教材・教示法を開発することにより、当初の目的とは異なる形ではあるが人と人とをコビキタスに結ぶ情報検索技術の利活用に新たな研究目標を据えることとした。

3. 研究の方法

(1) 研究開始当初の目的に対する実施

研究開始当初の目的に対しては、東京農工大学中川正樹研究室のメンバーと共同で、古典籍の解読支援のための Web アプリケーションの開発を行った。この Web アプリケーションは、過去の研究において申請者が開発した古文書解読支援のための Web 字形検索システム「MOJIZO」をベースに、中川研究室で開発された深層学習による手法を搭載したものであり、画像データを入力として解読結果の候補となる文字列を出力する end-to-end のサービスを提供する。

前述の MOJIZO では、ユーザーインターフェースを提供する Web サーバーと、検索機能を提供するサーバー（以下、コアサーバー）を独立させた形で設計・実装していた。この設計・実装は、サービスを提供する機関が独自性のあるユーザーインターフェースを実装することを可能にしていた。また、コアサーバーを DMZ に設置する必要がないため、Web サービスの提供に関する各機関のポリシーに影響を受ける / 及ぼすことなくサービスを提供できるという利点も有していた。このように機能ごとに独立 / 階層化された設計・実装は、機関を超えた共同研究の実施においても有効に機能した。本研究では、コアサーバー内においてもさらに機能ごとの独立 / 階層化を進め、研究・開発の成果を柔軟にサービスに反映できる方法を採用している。また、階層間のプロトコルにも一定の自由度を設け、今後新たなサービスを提供する場合でもサーバー全体への変更が最小限になるように工夫を行った。ただし、当該研究の成果は解読結果候補という抽象化された検索結果の出力に留まることになった。階層化には、利用する技術の組合せを柔軟に変更できるという利点がある。その点を活かせなかったことについて、本研究課題が期間途中で目的変更を余儀なくされたことは残念であった。

また、信州大学の白井敬一郎氏との共同研究を通して、古文書字形 / 花押など毛筆で記された情報の検索サービスを高精度化するための画像処理手法の改善にも取り組んだ。この研究では、毛筆で記述された文字 / 図形に残る運筆の形跡を画像処理で検出・強調し、経年変化、汚損などによって生じる変形 / ノイズに対する頑健性の獲得を目指したものであった。さらに、利用者によるアノテーションも併用しつつ、字形検索サービスがキーとなる画像内の情報を適切に利用することで検索精度全体を向上させるための技術の試作も行った。提案手法は評価実験を通して一定の有効性を示した。

(2) 修正後の目的に対する実施

一方で、修正後の目的に対しては、上記の検索技術で重要な役割を担った画像処理の技術を教材化し、学習者 / 学修者が「データを読む、説明する、扱う」といった数理・データサイエンス・AI の基本的な活用法に注目した学びの機会を得られる教材・教示法の実現を目指した。この教材・教示法は、情報科学に限らず幅広い学問領域を志す学習者 / 学修者への提供を目指すため、特定のプログラミング環境 / 言語は必ずしも適切ではない。一方で、学習者 / 学修者がすでに実施している学びとの連携は、既学修の内容を活かしたスムーズな理解、既学修内容の価値を体感する相互作用、の両面で有益であり、学習者 / 学修者のスキルをゼロベースに設定した学びに拘る必然性は薄い。そこで、本研究課題では代表的な表計算ソフトウェアである Microsoft Excel 上で動作するスプレッドシート型アプリを主要素とした教材、およびそれを活用した教示表を試作した。

実データとしては、学習者 / 学修者あるいは学びを提供する教師が用意した任意のデジタル画像ファイルを読み込めるようにした。ファイルを開く際には、Windows における標準的な「ファイルを開く」ダイアログを利用できるようにした。読み込まれた画像は、画素に見立てたセルに対して 1 画素 / セルで表示され、学習者 / 学修者は色（セルの塗りつぶしの機能を利用）だけでなく ARGB 形式の 16 進数で表された各画素の情報を視認できるようにした。以上の機能により、データを「読む」力を養う学びの実現を目指した。

また、別の 3 枚のシートには、画素ごとの R（赤）、G（緑）、B（青）それぞれの値を、数値および諧調画像として表示できるようにした。ここでは、数値として 10 進数での表現を採用し、前述の 16 進数による表現との対比を可能にすると共に、学習者 / 学修者による直感的な理解の支援を目指した。これらの機能を通して、学習者 / 学修者が「デジタルカラー画像における青い画素は、『青が強い』画素とは限らない。むしろ、『緑、赤が相対的に弱い』画素である。」といった「説明」を、自身の経験・体験に基づいて行うことのできる学びの実現を目指した。

さらに、別のシートでは、前述の R、G、B の各値を用いた簡単な計算式で 3 次元 1 次元の次元圧縮を行い、大津の 2 値化（判別分析法）を用いた領域抽出を経て注目画素 / 領域を切り出す簡単な画像処理 AI の機能を実装した。計算式としては、Excel のセルで動作する任意のもの（関数を含む）が利用可能だが、先の「説明」に基づき、B-R（あるいは B-G）のような簡単な引き算でも、ある程度の精度で青空を建物・地面・樹木などと区分できる事例を示すことができた。以上の機能により、コンピュータが持つ計算の機能を使ってデータを「扱う」ことを学ぶとともに、多数のデータを用意することで統計的な処理が可能となり、AI に求められる柔軟な条件分岐ができることを学ぶための教材および教示法の実現を目指した。

スプレッドシート型アプリには大津の 2 値化の動作を学ぶためのシートも用意した。このシ

ートでは、大きさが異なる 2 つの正規分布に対して平均値や中央値よりも適切な閾値を導出できることを視覚的に確認できるようにした。また、AVERAGEIF / COUNTIF などの基本的な関数が目的関数の実装に利用できることを示し、表計算ソフトウェアでの既習内容との関連性についても示した。さらに、一連の学びの内容を補足する PDF 形式の文書も教材の一部として同梱し、学習者 / 学修者が学びを一層深められるように工夫した。

このスプレッドシート型アプリについては、学生の学問的思考が多岐に渡るリベラルアーツ教育環境を採用した大学の基礎教育科目において授業実践を実施した。同実践では数十名を被験者とした予備的な評価実験も実施し、同アプリが AI 学修に対する学生のモチベーション向上に寄与する可能性が示唆された。被験者の人数を増やした本格的な実践研究については本研究課題の期間内に完了することができなかったが、今後の取り組みを通して効果を明らかにし、研究活動の有効性を示したいと考えている。

なお、教材・教法の研究に先立ち、研究フィールドになった基礎教育科目における教育上の課題を洗い出し、授業改善のための検討を実施するための取り組みも実施している。この取り組みについても、上記の教材・教法における研究活動の実施に資する知見・分析結果を得る機会になったものと考えている。

4. 研究成果

当初 / 修正後のそれぞれの目標に対する成果を整理する。前章「研究の方法」との間に内容の重複が生じている点についてご理解をいただくと幸いです。

(1) 当初の目的に対する成果

機械学習の応用と Web アプリケーション化

東京農工大学中川正樹研究室との共同研究により、深層学習による古文書画像の解読支援システムを Web アプリケーションとして実装し、情報検索のためのユーザーインターフェースとしての可能性を検証した。主な研究発表は次の通りである。

[1] 末代誠仁, リー・トゥアン・ナム, グエン・コング・カー, 中川正樹, 山本和明: 階層化された情報システムのためのくずし字解読機能の試作, 日本情報考古学会 第 42 回大会 講演論文集, Vol. 22, pp.11-15, 2019, 査読あり.

古代木簡解読支援のための画像処理の改善

信州大学白井敬一郎氏との共同研究により、画像中に残された運筆情報を抽出することで字形画像検索の精度を高めるための研究を行った。主な研究発表は次の通りである。

[2] 末代 誠仁, 白井 啓一郎: 古文書字形画像検索のための画像処理手法の実装, 日本情報考古学会 第 43 回大会 講演論文集, Vol. 23, pp.82-87, 2020, 査読あり.

歴史学に関連する研究で得た知見を用いて、著者の一人として文化情報学に関する事典を執筆した。成果の詳細は次のとおりである。

[3] 末代 誠仁: 文化情報学事典 (村上征勝 他著, 申請者の担当箇所は「A3-11 古文書に対する文字認識技術」の章), 勉誠出版, ISBN978-4-585-20071-0.

歴史学を含む人文学領域への AI の展開について、招待講演として知見を述べた。成果の詳細は次の通りである。

[4] 末代 誠仁: 情報処理学会「人文科学とコンピュータ研究会」における AI 研究の動向, 名古屋大学「デジタル人文社会科学研究推進センター」設置記念シンポジウム「AI で切り拓く人文学の未来」, 2024, 招待講演.

(2) 修正後の目的に対する実施

歴史情報および AI に関する研究成果の教育利用

古文書画像をキーとした情報検索技術の開発を通して得た技術・知識を AI リテラシー教育に活用するための教材・教法を開発し、申請者が本務先とする大学をフィールドとした実証実験を通して教育面での可能性を検証した。主な研究成果は次のとおりである。

[5] 末代 誠仁, 出井 智子: AI 学修のモチベーションを高めるための副教材を用いた予備実験, 情報処理学会 コンピュータと教育研究会 研究報告, Vol. 2023-CE-171(11), pp. 1-6, 2023, 査読なし.

以上。

5. 主な発表論文等

〔雑誌論文〕 計5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 3件）

1. 著者名 未代 誠仁, 出井 智子	4. 巻 2023-CE-171(11)
2. 論文標題 AI学修のモチベーションを高めるための副教材を用いた予備実験	5. 発行年 2023年
3. 雑誌名 情報処理学会 コンピュータと教育研究会研究報告	6. 最初と最後の頁 1-6
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 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 未代誠仁	4. 巻 2021-CH-127(7)
2. 論文標題 リベラルアーツ教育環境におけるコロナ禍での情報教育の一実践	5. 発行年 2021年
3. 雑誌名 情報処理学会 人文科学とコンピュータ研究報告	6. 最初と最後の頁 1-2
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 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 未代誠仁	4. 巻 2021-CE-161(11)
2. 論文標題 コロナ禍におけるコンピュータリテラシー科目での取り組みと考察	5. 発行年 2021年
3. 雑誌名 情報処理学会 コンピュータと教育研究会研究報告	6. 最初と最後の頁 1-7
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 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 未代 誠仁, 白井 啓一郎	4. 巻 22
2. 論文標題 古文書字形画像検索のための画像処理手法の実装	5. 発行年 2020年
3. 雑誌名 日本情報考古学会 第43回大会 講演論文集	6. 最初と最後の頁 82-87
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 未代誠仁, リー・トゥアン・ナム, グエン・コング・カー, 中川正樹, 山本和明	4. 巻 22
2. 論文標題 階層化された情報システムのためのくずし字解読機能の試作	5. 発行年 2019年
3. 雑誌名 日本情報考古学会講演論文集	6. 最初と最後の頁 11-15
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7件 (うち招待講演 1件 / うち国際学会 1件)

1. 発表者名 未代誠仁
2. 発表標題 情報処理学会「人文科学とコンピュータ研究会」におけるAI研究の動向
3. 学会等名 名古屋大学「デジタル人文社会科学研究推進センター」設置記念シンポジウム「AIで切り拓く人文学の未来」(招待講演)
4. 発表年 2024年

1. 発表者名 未代誠仁
2. 発表標題 AI学修のモチベーションを高めるための副教材を用いた予備実験
3. 学会等名 情報処理学会 コンピュータと教育研究会 第171回研究発表会
4. 発表年 2023年

1. 発表者名 Akihito Kitadai
2. 発表標題 Implementation and Teaching Plan of Spreadsheet Application for Basic Education of Artificial Intelligence
3. 学会等名 IFIP WCCE 2022: World Conference on Computers in Education (WCCE 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 未代誠仁
2. 発表標題 リベラルアーツ教育環境におけるコロナ禍での情報教育の一実践
3. 学会等名 情報処理学会 人文科学とコンピュータ研究会第127回研究発表会
4. 発表年 2021年

1. 発表者名 未代誠仁
2. 発表標題 コロナ禍におけるコンピュタリテラシー科目での取り組みと考察
3. 学会等名 情報処理学会 コンピュータと教育研究会第161回研究発表会
4. 発表年 2021年

1. 発表者名 未代誠仁
2. 発表標題 古文書字形画像検索のための画像処理手法の実装
3. 学会等名 日本情報考古学会 第43回大会 (COVID-19対応のため紙上開催)
4. 発表年 2020年

1. 発表者名 未代誠仁
2. 発表標題 階層化された情報システムのためのくずし字解読機能の試作
3. 学会等名 日本情報考古学会 第42回 大会
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 村上征勝 他（研究代表者が一部執筆を担当）	4. 発行年 2020年
2. 出版社 勉誠出版	5. 総ページ数 11
3. 書名 文化情報学事典（担当箇所は「A3-11 古文書に対する文字認識技術」の章）	

〔産業財産権〕

〔その他〕

漢字データの横断検索システムを開発へ https://www.asahi.com/articles/ASM3N3R4PM3NPOMB006.html

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
---------------------------	-----------------------	----

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
---------	---------