様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　１２　日現在

機関番号：　３２６０５
研究種目：　基盤研究(C)（一般）
研究期間：　2015～2017
課題番号：　１５Ｋ０２８４１

研究課題名（和文）古文書字形の機関横断的デジタルアーカイブの拡充・活用を支援する情報技術

研究課題名（英文）Crossover information search technologies to extend utilizations of digital archives of historical character patterns provided by multiple research organizations.

研究代表者

来代　誠仁（Kitadai, Akihito）

桜美林大学・総合科学系・准教授

研究者番号：００４０１４５６

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：本研究課題では、複数の研究機関が管理する古文書字形デジタルアーカイブの横断検索を実現する情報検索技術を設計・実装した。当該技術は、奈良文化財研究所のWebサイトで公開されており、利用者は任意の字形画像を検索キーとして入力できる。検索対象は、同研究所の木簡庫と、東京大学史料編纂所の電子くずし字字典デジタルアーカイブである。約19カ月の評価実験において、当検索には936,932個の画像キーが入力された。この数はテキストによる検索手法に入力された検索キー数の2倍以上となり、本研究課題で実現した検索技術が古文書デジタルアーカイブの情報検索に対する新しいニーズを開拓した可能性を示した。

研究成果の概要（英文）：In this research, I have designed and implemented information technologies for crossover searching of digital archives of historical character images. The technologies, accepting character pattern images as the search key, have been provided by a web site of Nara National Research Institute for Cultural Properties. The targets of the searching are the digital archives of the institute and Historiographical Institute, The University of Tokyo. In our evaluation experiment, 936,932 image keys have been used for the searching in about 19 months. The number is about two times as large as that of the keys of the search method using character codes as the keys. The results show that the new technologies satisfy the needs for information searching of the digital archives.

研究分野：　情報学

キーワード：　古文書デジタルアーカイブ　情報検索　パターンマッチング

１．研究開始当初の背景

当研究課題が開始した当初において、古文書デジタルアーカイブは既にボーンデジタルの時代を迎えていた。すなわち、それまでのガラス乾板、フィルムといったアナログ媒体上の記録をデジタル化するに留まらず、現物の古文書を直接デジタル情報として記録・保存することがアーカイビングの自然な流れとして定着しつつあった。記録された情報はデジタル機器によって使用されることを前提としていることになり、コンピュータネットワークを用いた古文書のアプリケーションには必然的に大きな期待が集まっていた。

情報のデジタル化は、記録媒体の物理サイズに影響を受けにくい情報の管理方法である。図書館１件分の情報が片手に乗る記録媒体に収まるという事実には大きなメリットがある。しかし、1 点の収録遺物が「どこ」にあるのかを知るために、情報検索技術への依存は避けられない。「西館 3F の書棚の下段真ん中くらい」といった空間を意識した管理はデジタル化された情報に対して有益とはいえない。

当研究開始当初の時点において、国内で公開されている多くの古文書デジタルアーカイブには、それぞれ開発者の工夫が凝らされた情報検索技術が搭載されていた。しかし、その大部分は専門家が１点ずつの収録遺物に付与したメタデータと呼ばれる情報をデータベースマネージメントシステム（DBMS）の標準機能で参照するという既存技術の上に成り立っていた。そのため、検索技術の有用性は「利用者とメタデータフォーマットの相性」に依存していた。このことは、デジタルアーカイブに収録された情報を利用者に幅広く結びつけ、古文書の英知を後世の社会に活かすという**古文書デジタルアーカイブの大きな目標を果たす上での壁**といえる問題であった。

以上のことから、①多様な古文書デジタルアーカイブに適用可能で、かつ利用者がメタデータフォーマットに特別な意識を持つ必要がない（シームレスな）情報検索技術を実装すること、②その情報検索技術によって古文書デジタルアーカイブの利用が活性化され、古文書が持つ情報と利用者を強く結びつけることができることを明らかにすること、の２点が、古文書デジタルアーカイブに関する研究の大きな課題であると申請者は考えるようになった。

２．研究の目的

前述の背景を踏まえて、本研究課題においては、様々な古文書から抜き出した字形画像デジタルアーカイブを検索対象として、汎用フォーマットであるデジタル画像を検索キーとした情報検索技術を実装し、インターネットを通して幅広く公開することによって、前述の**メタデータフォーマットによる壁の影響を緩和**できること、すなわち古文書デジタルアーカイブの利用を活性化し、知識を利用者と強く結びつけられることを明らかにすることを目標と定めた。この目標を達成することは、すなわち(a)メタデータおよびDBMS の標準機能に基づく既存検索技術には技術的発展の余地が残っていること、(b)古文書デジタルアーカイブには広く公開されるに至っていない価値がまだ残っていること､の２点を明らかにすることでもあった。

３．研究の方法

研究目標を達成するためには、まず字形画像をキーとした情報検索技術を Web サーバ上に実装し、多くの利用者に公開できる形に仕上げる必要があった。申請者は、既にスタンドアロン方式（利用者の PC にインストールする形態）の字形検索機能を、古代木簡解読支援ソフトウェア「Mokkanshop」に実装することに成功していた。しかし、Web サーバ上での実装においては、マルチユーザに対応した並列処理、デジタルアーカイブへの排他アクセス制御、**24 時間 365 日**の運用に耐える安定性などを確保する必要があった。また、当該検索技術の有用性が示された後での継続的な運用を見据えると、サーバコンピュータへの負荷を現実的な範囲に抑制することも重要な課題であった。このような実装が可能かどうかは研究レベルにおいても明らかではなかったが、申請者はスタンドアロン方式を前提とした既存の実装を全面的に見直し、必要となる機能・性能を達成することに成功した。なお、現時点に至るまで、当実装の不備を原因とした Web サービスの不具合は発生していないことを申し添えておく。

ただし、前述の Mokkanshop に実装していた処理のうち、キーとなる画像のノイズ除去を行う画像処理技術については、Web サーバへの実装が適切ではないと判断した。画像処理は、パラメタの変更に対して処理結果となる画像を随時更新する「フィードバック」が必要である。しかし、Web サーバに画像処理を実装した場合、クライアント（利用者側コンピュータ）とサーバとの間でフィードバックのための通信が頻発してしまう。インターネットを利用した Web サービスにおいては、利用者が従量課金回線を利用している可能性を考慮する必要があるため、データ通信量削減への配慮は不可欠である。そこで、画像処理機能についてはインターネット利用時の主なクライアントになりつつある iPhone 用アプリとして実装し、利用者に配布する方針を採ることにした。

次に、検索対象となる古文書字形画像デジタルアーカイブとの連携を実現する必要があった。これについては、申請者が研究分担者として参加している別の科研費などで実現された奈良文化財研究所、東京大学史料編纂所の木簡字典（現：木簡庫）、および電子くずし字字典データベースの２つのデジタル

アーカイブを検索対象とすることで、当課題の研究費の効果的な利用に配慮した。2 つのデジタルアーカイブには、テキストによる横断検索機能（1 個のテキストキーで 2 つのデジタルアーカイブを同時検索する機能）が既に実装されていたが、字形画像キーによる検索も同様に横断検索を行うように実装を行った。

最後に、利用活性化の評価方法については、奈良文化財研究所の協力を得て、同所の Web サイト上に、字形画像キーによる横断検索を提供する Web アプリを設置してもらい、前述のテキストキーによる検索と並行した利用件数の記録を行ってもらうことができた。テキストキーによる検索は、2 つのデジタルアーカイブが持つ「文字画像の字種情報」に関するメタデータを参照することで実現されている。この検索および字形画像キーによる検索を同一の Web サイト上で評価することにより、活性化の評価に客観性を与え、また相互に与える影響を評価することにもつながると考えた。

４．研究成果

当研究課題の遂行を通して実現した情報検索技術を搭載した、字形画像キーによる古文書字形画像デジタルアーカイブ検索サービス「MOJIZO」の全体構成を図 1 に示す。

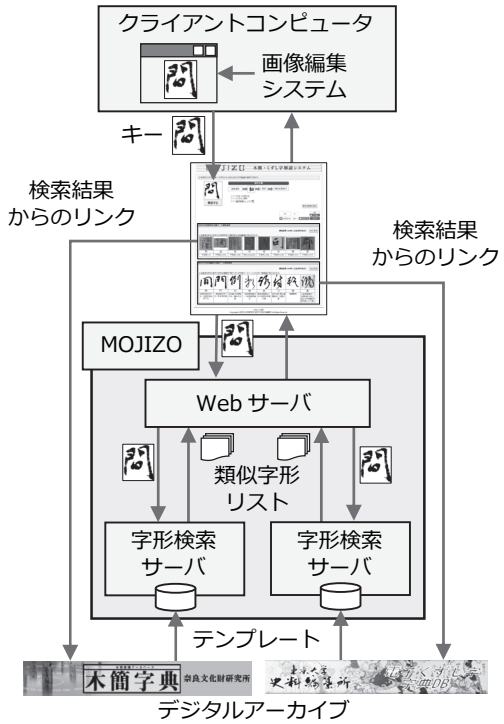

図 1. 字形画像キーによる検索サービス「MOJIZO」の全体構成

当研究課題で実現した字形検索技術の実装となる「字形検索サーバ」を含めて、MOJIZO の機能は Web サーバ上で動作する。利用者は、Web ブラウザを搭載した任意のコンピュータをクライアントとして利用可能である。

次に、申請者が iPhone 用画像処理アプリとして実装した「MOJIZOkin」について図 2 に示す。

図 2. iPhone 用画像処理アプリ「MOJIZOkin」

画像処理は、古文書の状態によって適切なものを選択する必要がある。これについて、PC 用アプリケーションソフトウェアについては、利用者にある程度の選択が用意されている可能性が高い。しかし、現在の利用者にとってインターネット利用時の主たるクライアントであるスマートフォンにおいては、適切な画像処理の手段がないのが現状である。iPhone 用アプリの提供を通してクライアントコンピュータに対する制限を緩和することは、字形画像キーによる検索への評価を現実的な環境で実施する上で重要だと考える。

奈良文化財研究所の Web サイトで計測した MOJIZO（字形画像キー）、およびテキストキーによる検索件数は図 3 の通りである。ただし、MOJIZO の公開は平成 27 年 3 月のため、H27 年度分として表記されているのは 1 カ月未満分である。また、平成 29 年度の 10 月以降については現在調査中のため、H29 年度分としては前半 6 か月のみを表記している。この結果から、MOJIZO はテキストキーによる検索を上回るペースで利用されており、デジタルアーカイブの利用活性化を実現した可能

性が極めて高いこと、および、既存のテキストキーによる検索機能の利用件数に悪影響を与えることなく、デジタルアーカイブに対する新たなニーズを開拓した可能性が高いことが示された。

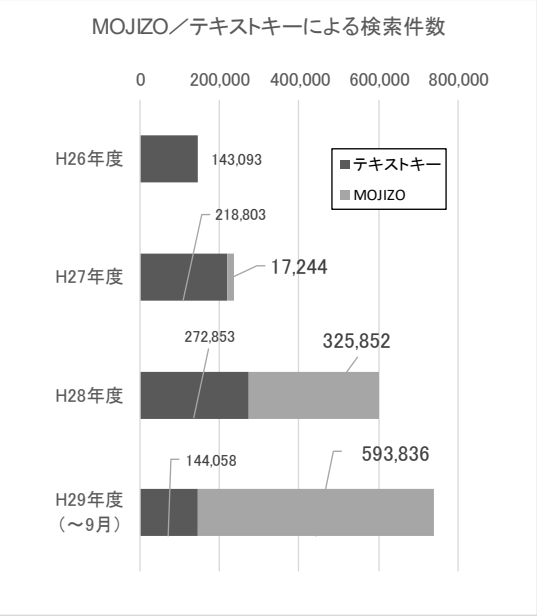

図 3. デジタルアーカイブごとの検索件数

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　7 件）

① 未代誠仁、高田祐一、井上幸、方国花、馬場基、渡辺晃宏、井上聡，字形画像をキーとした情報検索による古文書デジタルアーカイブ活用への効果，情報処理学会論文誌，査読有，Vol. 59-2，2017，pp. 351-359.
② A. Kitadai, M. Inoue, Y. Tanaka, G. Fang, H. Baba, A. Watanabe and S. Inoue, Technologies and Improvements of Image Search Service for Handwritten Character Patterns on Japanese Historical Documents, Proceedings of the 14th International Conference on Document Analysis and Recognition (ICDAR 2017)，査読有，Vol. 1，2017，pp. 1180-1185.
③ 未代誠仁，字形検索サービスにおける文字認識技術の活用，第3回日本語の歴史的典籍国際研究集会 発表要旨・発表資料集（国文学研究資料館），査読無，Vol. 1，2017，pp. 11-12 and 56-59.
④ 未代誠仁，文字画像検索システム MOJIZO について，情報処理学会　人文科学研究会，第 115 回研究会 予稿集，査読無，Vol. 2017-CH-115(7)，2017，pp. 1-2.
⑤ 未代誠仁，井上幸，高田祐一，方国花，馬場基，渡辺晃宏，井上聡，木簡およびくずし字のデジタルアーカイブを文字画像で検索するサービスの実装，情報処理学会 人文科学とコンピュータシンポジウム「じんもんこん 2016」論文集，査読有，Vol. 1，2016，pp. 19-24.
⑥ A. Kitadai, Y. Takata, M. Inoue, G. Fang, H. Baba, A. Watanabe, S. Inoue, A Web Based Service to Retrieve Handwritten Character Pattern Images on Japanese Historical Documents, Proc. 6th Conf. Japan Association for Digital Humanities (JADH 2016)，査読有，Vol. 1，2016，p. 57.
⑦ 未代 誠仁，馬場 基，渡辺 晃宏，井上聡，久留島 典子，中川 正樹，古文書字形デジタルアーカイブのための検索システムの試作，じんもんこん 2015 論文集，査読有，Vol. 2015，2015，pp. 9-15.

〔学会発表〕（計　12 件）

① 未代誠仁，現代のロゼッタ・ストーンができた！古文書の読めない文字を読み解くアプリ，つくばサイエンスエッジ（主催：つくば ScienceEdge2018 実行委員会、後援：茨城県、つくば市、文部科学省、JST 他），2018 年 3 月 24 日，つくば国際会議場.
② 未代誠仁，「第 21 回 PRMU アルゴリズムコンテスト CH 賞受賞式」パネルディスカッション（パネリストとして登壇），情報処理学会 人文科学研究会 第 116 回研究発表会，2018 年 1 月 27 日，函館コミュニティプラザ G スクエア.
③ A. Kitadai, M. Inoue, Y. Tanaka, G. Fang, H. Baba, A. Watanabe and S. Inoue, Technologies and Improvements of Image Search Service for Handwritten Character Patterns on Japanese Historical Documents, The 14th International Conference on Document Analysis and Recognition (ICDAR 2017)，2017 年 11 月 15 日，Kyoto Terrsa.
④ 未代誠仁，文字画像検索システム MOJIZO について，情報処理学会 人文科学研究会 第 115 回研究発表会（主催者による企画セッション内），2017 年 8 月 4 日，東京大学史料編纂所.
⑤ 未代誠仁，字形検索サービスにおける文字認識技術の活用，国文学研究資料館 第 3 回日本語の歴史的典籍国際研究集会（招待講演），2017 年 7 月 28 日，国文学研究資料館.
⑥ 未代誠仁，字形画像による情報検索技術の可能性と課題，東京大学史料編纂所 公開研究会「歴史学と情報―研究資源の新たな利活用に向けて」（招待講演），2017 年 6 月 2 日，東京大学史料編纂所.
⑦ 未代誠仁，デジタルアーカイブの利活

用を促進する情報検索技術の研究を通して感じた課題（兼、テーマセッション『:「人文科学とコンピュータ」分野が一層発展するための課題は何か？』パネリスト），情報処理学会 人文科学研究会 第114回研究発表会，2017年5月13日，龍谷大学（京都）．

⑧ 未代誠仁，歴史学の情報 part3 ～読めない文字への挑戦～，情報処理学会 IPSJ-One 2017（招待講演），2017年3月18日，名古屋大学 豊田講堂．

⑨ 未代誠仁，木簡およびくずし字のデジタルアーカイブを文字画像で検索するサービスの実装，情報処理学会 人文科学とコンピュータシンポジウム「じんもんこん2016」，2016年12月10日，国立国語研究所（立川）．

⑩ A. Kitadai, A Web Based Service to Retrieve Handwritten Character Pattern Images on Japanese Historical Documents, 6th Conf. Japan Association for Digital Humanities (JADH 2016), Sept. 13, 2016, The university of Tokyo.

⑪ 未代誠仁，古文書字形の研究成果を公開するための技術，Workshop: “Management of Japanese Character Information and its Application” in 6th Conf. Japan Association for Digital Humanities (JADH 2016, 公開セッション), 2016年9月12日，東京大学 福武ホール．

⑫ 未代誠仁，デジタル技術による分析と経験知の融合にむけて―文字の数値的分析技術から見た可能性，シンポジウム「字体と漢字情報」―HNG 公開 10周年記念―，2015年11月21日，国立国語研究所．

〔図書〕（計 2件）

① 渡辺晃宏，未代誠仁，日本工業出版，画像ラボ2017年10月号「文字の世界を開く 文字画像データベースMOJIZOの開発」，2017, pp. 22-29.

② 未代誠仁，勉誠出版，デジタル技術による分析と経験値の融合にむけて（高田智一、他 編「漢字字體史研究」の一章として），2016, pp. 331-346.

〔産業財産権〕

○出願状況（計 0件）

○取得状況（計 0件）

〔その他〕
ホームページ等

① MOJIZO,
http://mojizo.nabunken.go.jp/

② MOJIZOkin（app store内）
https://itunes.apple.com/jp/app/mojizokin/id1211838518?mt=8

６．研究組織

(1)研究代表者

未代 誠仁（KITADAI, Akihito）
桜美林大学・総合科学系・准教授
研究者番号：00401456

(2)研究分担者

該当なし

(3)連携研究者

該当なし

(4)研究協力者

該当なし