情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

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

来代 誠仁[1,a)] 高田 祐一[2] 井上 幸[3] 方 国花[2] 馬場 基[2] 渡辺 晃宏[2] 井上 聡[4]

受付日 2017年5月9日, 採録日 2017年11月7日

概要：古文書の研究者にとって，古文書デジタルアーカイブの活用を促すことは重要な課題である．本論文では，字形画像をキーとした横断検索技術による古文書 Web デジタルアーカイブ活用への効果について述べる．字種は文書に対する現実的な検索キーの 1 つである．しかし，古文書において字形との対応は必ずしも確定しない．この課題を解決するために，私たちは字形画像をキーとした古文書 Web デジタルアーカイブの横断検索を実装した．5 カ月間の実験で入力されたキー数は合計で 200,000 件を超えた．これは字種による横断検索の件数と比較しても十分に大きい．また，私たちは古文書解読の専門家による評価実験を実施した．専門家は，使いなれた画像処理ソフトウェアを搭載した PC もしくは筆者らが作成した画像処理ソフトウェアを搭載した iPod Touch，またはその両方を使用した．「検索結果にキーと類似した画像が含まれるか」という旨の質問に対しては，すべての専門家が肯定的な回答を示した．検索精度と使い勝手の向上，および字形テンプレートの整備を通した活用のさらなる促進は今後の課題である．

キーワード：デジタルアーカイブ，古文書，情報検索，字形画像検索

Activating Impacts on Digital Archives of Historical Documents by Information Search with Character Pattern Image Keys

Akihito Kitadai[1,a)] Yuichi Takata[2] Miyuki Inoue[3] Guohua Fang[2]
Hajime Baba[2] Akihiro Watanabe[2] Satoshi Inoue[4]

Received: May 9, 2017, Accepted: November 7, 2017

Abstract: Increasing the uses of digital archives of historical documents is an important aim for researchers of the documents. In this paper, we show the effects of employing character pattern image keys for crossover search of our Web-based digital archives of the historical documents. Character codes are reasonable keys for the search. However, definitions of relationships between the codes and character shapes on the historical documents are ongoing research activities of history and archaeology. Therefore, we added a function to receive character pattern images as the keys. It creates the alternative relationships between the keys and the character pattern images of the digital archives. In a 5-month experiment, the total number of the image keys for the search was over 200,000, it was comparable to the number by character code keys in the same term. We also conducted an evaluation experiment by expert readers of historical documents. Each reader used PC with favorite image processing software, or "iPod Touch" with our image processing software, or the both. All readers returned positive responses to the question that "Does the results of the search by character image keys contain character images like the keys?". Improving the accuracy and the usability, and refining the templates of the search are our future work to obtain more uses.

Keywords: Digital archives, Historical documents, Information search, Character pattern image search

1 桜美林大学
J. F. Oberlin University, Machida, Tokyo 194–0294, Japan
2 奈良文化財研究所
Nara National Research Institute for Cultural Properties, Nara 630–8577, Japan
3 東大阪大学
Higashiosaka College, Higashiosaka, Osaka 577–8567, Japan
4 東京大学史料編纂所
Historiographical Institute The University of Tokyo, Bunkyo, Tokyo 113–0033, Japan
a) a.kitadai@gmail.com

情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

1. はじめに

古文書に関する研究成果を管理，再利用する手段として，デジタルアーカイビングに対する注目が集まっている．古文書および各種研究成果を書棚，保管庫などの物理的サイズに依存することなく管理し，必要な情報を短時間かつ選択的に参照できるデジタルアーカイブは，現代の古文書研究にとって不可欠な存在になりつつある．

一覧表示が困難となる量の情報から必要なものを選択するには，適切な情報検索技術が必要である．少なくとも，利用者にとっては情報検索技術を用いずにデジタルアーカイブを利用することは非現実的といえる．古文書デジタルアーカイブの有用性が研究成果の活用にあると考えるとき，多くの利用者のニーズに沿った情報検索技術の提供は重要な課題となる．

Coreに代表されるメタデータ記述の標準化は，様々な文化財を収録したデジタルアーカイブの使い勝手を共通化し，さらに複数のデジタルアーカイブの横断検索に道を開く高い有用性をもたらした[1]．現在，様々な文化財を横断的に検索して利用者に提供するための世界的な取り組みがすでに存在している[2], [3], [4]．その一方で，検索対象となるコンテンツの種類が限定できる場合は，該当するコンテンツに特有のメタデータを適切に扱うことができる情報検索技術によって高い有用性が実現する可能性もある．

筆者らの所属する研究機関のWebサイトでは，古文書に関する研究成果を扱ういくつかのWebデジタルアーカイブを提供している．それらの中に，古代木簡を収録対象とする「木簡字典」と，平安時代後期から近世初頭までの和紙文書を収録対象とする「電子くずし字字典データベース」がある[5], [6]．2つのデジタルアーカイブは，その名前が示すとおり古文書の文字に関する研究成果を含んでいる．具体的には，古文書から切り出した1文字分の字形画像，および字種を格納するメタデータの書式を有している．筆者らはこの共通性に着目し，1文字の字種をキーとする字形画像の横断検索サービスを提供してきた[7]．この横断検索サービスでは，デジタルアーカイブ内で字形画像と他データとの間に張られたリンクを利用することで，古文書デジタルアーカイブそのものの横断検索も実現している．

字種による字形画像および古文書デジタルアーカイブの検索は，横断検索への発展性を含めて，古文書の文字に関する研究成果をWebで公開する際の現実的な情報検索技術の1つといえる．ただし，古文書に記された字形と字種との関係が現在進行形の研究課題であることには注意が必要である．字種が確定していない字形画像は多数存在し，字種という分類が時代を超えて利用できるのかという点も検証の最中である．以上のことは，古文書の字形画像に対して，字種をキーとした情報検索だけでは対応しきれない利用者のニーズが存在しうることを示唆するものと考えられる．

字種に代わるキーとしては，利用者が用意した字形画像が考えられる．画像を対象としたパターンマッチング技術を用いてキーをデジタルアーカイブの字形画像と対応付けることで，理論的には情報検索が可能となる．画像情報のパターンマッチング技術を日本語の古文書デジタルアーカイブに応用した近年の研究には次のようなものがある．Panichkriangkraiらは，古典籍へのメタデータ付与を支援する文書解析システムを提案，実装した[8]．寺沢らは，古文書に記された任意長の文字列に対するワードスポッティング技術を実現し[9]，デジタルアーカイブ内の古文書間で類似性の高い部分を対応付けることができるWebアプリケーションを公開した[10]．早坂らは，変体仮名に対して深層学習による識別器を作成し，Webアプリケーションとして公開した[11]．北本らは，古典籍の画像に文字の座標情報と字種を付与し，パターン認識の研究に利用可能な形で公開する取り組みを行っている[12]．筆者らの研究グループにおいても，断片化した古文書へのメタデータ付与を支援するシステムを実現している[13]．ただし，このような国内の研究，あるいは海外の研究においても，字形画像をキーとした検索技術が，古文書のデジタルアーカイブの活用に与える効果は明らかにされていない．筆者らは，これまで研究を行ってきた古文書字形を対象としたパターンマッチング技術[14]を応用することで，字形画像をキーとした古文書Webデジタルアーカイブの横断検索を提供するWebアプリケーション「MOJIZO」を構築し，奈良文化財研究所のWebサイトにおいて公開した[15]．また，キーとなる字形画像の作成，編集を行うiPhone/iPod touch用の画像処理アプリ「MOJIZOkin」を構築し，App Storeで公開した[16]．本論文では，MOJIZOおよびMOJIZOkinの構築に用いた技術について述べるとともに，MOJIZOの利用状況，および古文書解読の専門家による評価実験の結果を示し，字形画像をキーとした検索技術がデジタルアーカイブの活用に有用であることを明らかにする．

2. 検索対象となる古文書デジタルアーカイブ

この章では，本論文で述べる横断検索の対象となる「木簡字典」と「電子くずし字字典データベース」の2つの古文書デジタルアーカイブ，および，字種をキーとした横断検索について述べる．

2.1 木簡字典

木簡字典は，奈良文化財研究所のWebサイトで公開している，古代木簡（図1）を収録対象としたデジタルアーカイブである．

古文書としての木簡は日本各地で40万点以上が見つかっているが，その約半数は平城宮跡とその周辺で発見された古代の木簡である．古代木簡のほとんどが遺跡のゴミ捨て

情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

図 1 古代木簡
Fig. 1 Mokkans in ancient times.

穴，溝，井戸跡などから出土していること，人為的な破壊の形跡があるもの/文字が記録された木片の表面を削り落とした削屑などが多数発見されていることなどから，古代木簡の主な用途は長期の保存を意図しない文書の作成であったと考えられている．このため，古代木簡には作成当時における人々の日々の営みが直接的に記録されている可能性が高い．ただし，古代木簡を解読するうえでは，先述の破壊痕，地中で受けた損傷，経年変化による変色・脱色などによる字形の損失が問題となる．専門家は，自然光/赤外光による墨痕の分析，記帳と呼ばれる観察記録の保存と共有などを通して解読作業を進めているが，ある程度の解読が進んだ古代木簡は一部にとどまる．

木簡字典には，解読作業にある程度の進捗がみられる約15,000点（表裏別）の古代木簡が収録されている．これらに対しては，釈文となるテキスト，木片の形状，大きさ，木材の種類，発見場所などがメタデータとして記録されている．また，古代木簡の全体画像に加えて，1文字分の字形画像も100,000点以上登録されている．全体画像および字形画像には，自然光（カラー，モノクロ）/赤外光によるデジタル画像，記帳をデジタル化した画像が含まれる．

木簡字典の使用時には，メタデータに対応するキーを用いた情報検索機能を利用する．たとえば，釈文に含まれる1文字以上のテキストをキーとして古代木簡の一部を一覧表示させ，さらにリンクを使って古代木簡の詳細な情報にアクセスすることができる（図 2）．木簡字典は，研究成果の一般公開に加えて，古代史の研究者が過去の研究成果を再利用することも大きな目的としている．難読字形に対しては，過去の類例を用いた検証が有効となるためである．このため，専門知識を要するキーによる詳細選択機能も提供している（図 3）．また，欠損が著しく形状の情報だけでは解読困難な字形画像も登録されている．

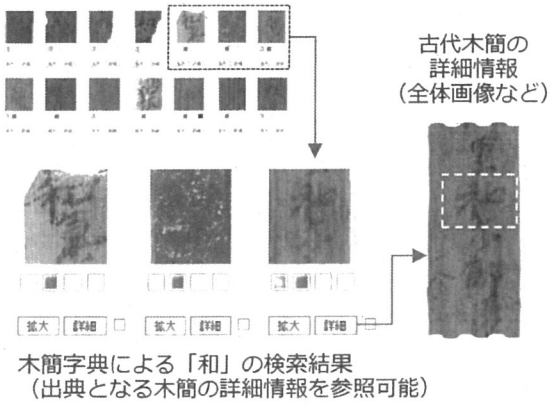

図 2 木簡字典を使った文字「和」の検索
Fig. 2 Document search of “和” on “木簡字典”.

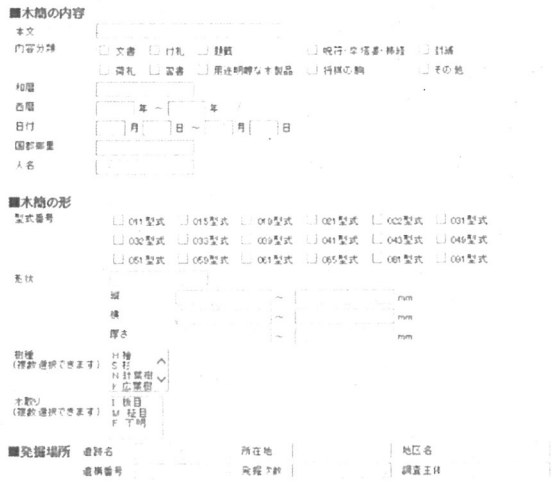

図 3 木簡字典の詳細検索画面
Fig. 3 Search refinement GUI of 木簡字典.

2.2 電子くずし字字典データベース

電子くずし字字典データベースは，東京大学史料編纂所のWebサイトで公開している，平安時代後期から近世初頭までの和紙文書（図 4）を収録対象としたデジタルアーカイブである．

東京大学史料編纂所では，和紙文書に記された様々な字形/字種を分析し，用途が類似する字種，字形が類似しやすい字種といった字形/字種の様々な関連性を調査してきた．電子くずし字字典データベースは，約24,000の字形画像に対して，字種に関する情報（コード，部首），字種間の関係，出典となる文書の名称/作成年/筆者，原本/影写本の区別などをメタデータとして付与したデジタルアーカイブとして公開された．その後，字形画像とメタデータの継続的な追加が行われている．

電子くずし字字典データベースで字種を指定した検索を行うと，用法/形状が類似しやすい字種へのリンクも取得できる（図 5）．また，部首/用途/時代など，字形/字種の分析結果を生かしたキーによる詳細検索も利用できる（図 6）．

情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

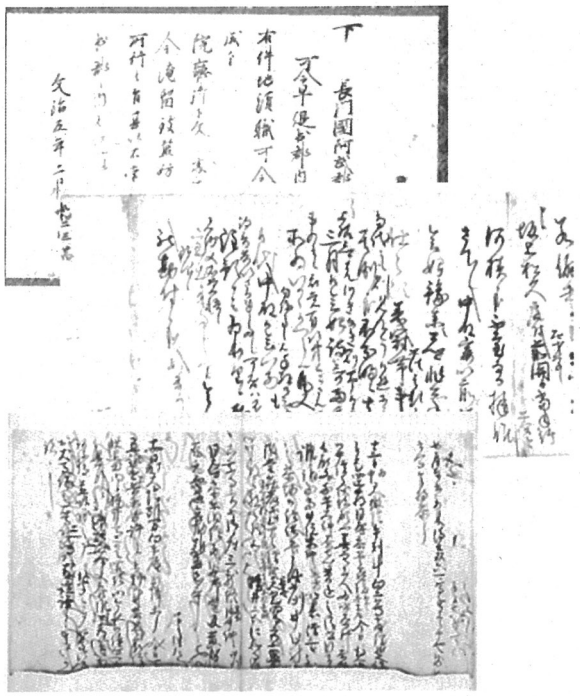

図 4 和紙文書
Fig. 4 Historical washi documents.

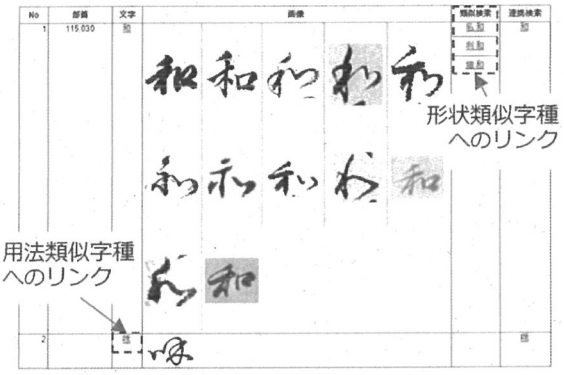

図 5 電子くずし字字典データベースを使った文字「和」の検索結果
Fig. 5 Document search results of “和” on “電子くずし字字典データベース”.

和紙文書の字形のくずし方は多様であり，解読に専門的な知識が必要となる．また，経年変化/破損などによる字形の損失，裏面の記述の映り込みなどによる難読字形も存在する．電子くずし字字典データベースの検索機能・コンテンツには，字形の多様性や様々な意図を記録し，難読字形の解読に役立つ，といった可能性が期待される．

2.3 横断検索

前述の 2 つのデジタルアーカイブには，コンテンツおよび研究上の特徴に起因する差異が存在するが，字形画像と字種の情報を有する点では共通している．この点を利用して，2 研究機関の Web サイトでは字種をキーとした横断検索を提供している．

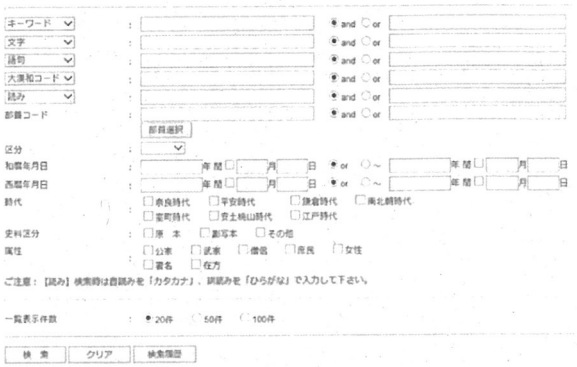

図 6 電子くずし字字典データベースの詳細検索画面
Fig. 6 Search refinement GUI of “電子くずし字字典データベース”.

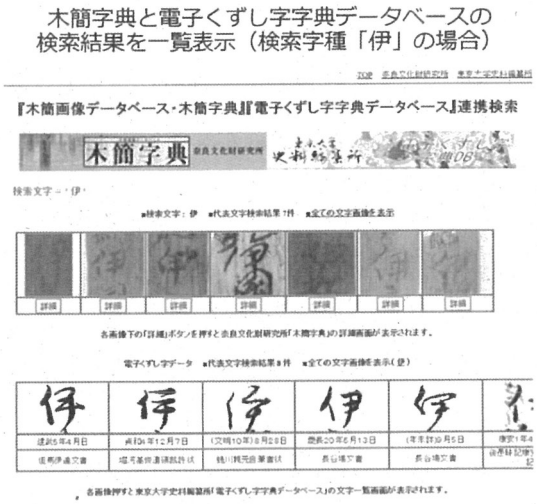

図 7 連携検索による字種「伊」の検索
Fig. 7 Crossover retrieval results of “連携検索” for character “伊”.

横断検索では，1 文字分の字種だけをキーとして入力することができる．検索結果は，それぞれのデジタルアーカイブに登録された字形画像の一覧として表示される（図 7）．個々の字形画像は出典となるデジタルアーカイブへのリンクになっており，利用者はリンクを通して字形画像の詳細な情報を得ることができる．横断検索の利用状況については後述する．

3. 字形画像をキーとした古文書デジタルアーカイブの検索技術

3.1 字種とは異なる検索キーの可能性

文字を用いた文書は，字種の列によって情報を保存・伝達する性質を持つ．したがって，字種による情報検索は，古文書デジタルアーカイブにとって現実的かつ必要な機能である．しかし，難読字形を多数含み，言語にも時代の差が存在しうる古文書のデジタルアーカイブにおいては，すべての字形を字

情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

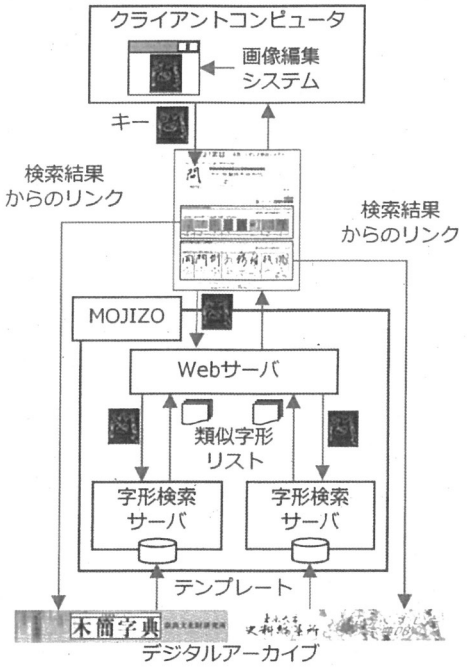

図 8 字形をキーとした検索のための構成図

Fig. 8 Composition figure to implement information search using character pattern image keys.

種によって管理することは困難である．このことは，字種による情報検索に制限が生じることを意味する．デジタルアーカイブの利用を促進するうえで解決すべき課題といえる．

本論文では，字種の代わりに字形画像をキーとした情報検索について述べるが，明らかにしたいのは，字種をキーとした検索が受ける制限を他のキーによって補いうるかどうか，という点である．字種をキーとした検索，および字種/字形画像以外をキーとした情報検索の有用性を否定するものではない．

本章では，筆者らが構築した字形画像をキーとする古文書 Web デジタルアーカイブ検索（以下，当検索）の技術について述べる．

3.2 情報検索のための構成

図 8 に，当検索を実現するために筆者らがとった構成を示す．

構成の中心となるのは，Web アプリケーション MOJIZO である．MOJIZO は，字形が黒で背景が白，あるいはそれに準ずる明暗のはっきりした字形画像をキーとして受け取ると，キーの形状を評価し，検索対象となるデジタルアーカイブに登録された類似性の高い字形画像を検索結果として表示する．利用者は，任意の画像処理システムを用いて MOJIZO に適した字形画像を作成・編集できる．なお，iPhone/iPod touch 用の画像処理アプリ MOJIZOkin については後述する．

MOJIZO が検索結果として表示する字形画像は，前述の

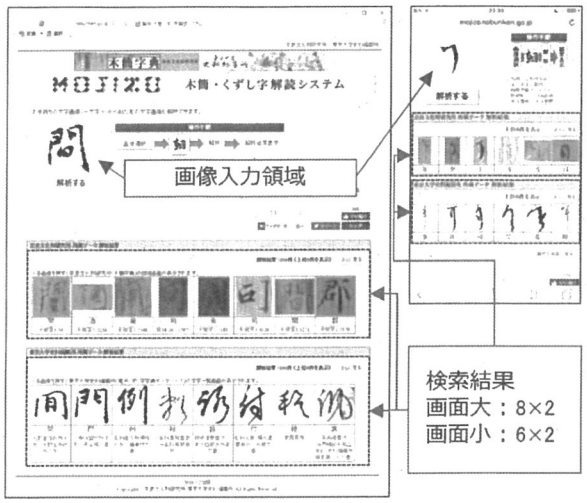

図 9 MOJIZO のユーザインタフェース

Fig. 9 User interface of MOJIZO.

字種による横断検索と同様に，出典となるデジタルアーカイブへのリンクを提供している．リンクによる移動後，利用者は各デジタルアーカイブの機能を利用して情報を閲覧することができる．

3.3 Web アプリケーション MOJIZO

MOJIZO では，ユーザインタフェースを提供する Web サーバと，字形評価処理を担当する字形検索サーバを分離し，同一/別々のコンピュータ上での動作を可能にした．これによって，MOJIZO を公開する研究機関では，ユーザインタフェースのデザイン変更，Web サーバへのネットワークポリシの適用などを柔軟に実施できる．字形検索サーバの数は，Web サーバの実装に合わせて任意に変更可能である．現在は，1 個の Web サーバ，および 2 つのデジタルアーカイブをそれぞれ担当する 2 個の字形検索サーバを組み合わせて運用を行っている．それぞれのサーバの処理はサーバサイドで担っており，クライアントとなるコンピュータへの負荷に配慮している．

Web サーバにキーとなる字形画像を入力する際には，画像入力領域をクリック/タップするか，同領域に画像をドラッグ&ドロップする（図 9）．多様な操作方法に対応することで，利用環境への制限緩和を目指している．Web サーバが検索結果として表示する字形画像の数は，当初はデジタルアーカイブごとに 8 個としていたが，現在はクライアントの画面が小さい場合には自動的に 6 個に変更する．また，結果表示画面全体のレイアウトも画面サイズに応じて変更する．ただし，いずれの場合も「さらに見る」のボタンを押すことで最大 100 個の字形画像が表示可能である．

字形検索サーバでは，線密度を用いた非線形正規化と勾配特徴の抽出を用いて，デジタルアーカイブの個々の画像をテンプレートとするパターンマッチングを行い，キーと

情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

なる字形画像との類似度をそれぞれ算出する．線密度による非線形正規化 [17], [18], [19] と勾配特徴 [20] は，それぞれ漢字圏における手書き文字認識で効果を示しており，筆者らも古文書の字形に対する有効性を確認している [21]．

テンプレートとしては，木簡字典からは約 650 の頻出字種に対応する 5,184 画像，電子くずし字字典データベースからは約 5,800 の頻出字種に対応する 23,548 画像を登録した．なお，ここでの字種数は現在の日本語に寄せたものであり，各時代における数とは必ずしも一致しない．テンプレート数の追加登録は筆者らの継続的な課題であり，パターンマッチングを用いる MOJIZO の性質に合わせて，字形単独での判読が可能な字形画像を中心にテンプレートの整備を進めている．

3.4 画像処理アプリ MOJIZOkin

クライアントとなるコンピュータの多様化，特に，利用者が多い PC とスマートフォンの存在は，デジタルアーカイブの活用を論じるうえで重要な検討課題である．

当検索の利用者は，用意した画像に合わせた任意の画像処理システムを利用してキーとなる字形画像を作成・編集できる．理想的には，Web アプリである MOJIZO が画像処理機能も一括提供するのが好ましいが，古文書/字形画像の多様性，Web アプリに対する利用者の慣れなどの要因を考慮すると，画像処理の手段に対する選択をクライアントと利用者に委ねることは現実的な選択と考える．ただし，スマートフォンのような小型のコンピュータについては，利用者が適当な画像処理アプリを探すことが現時点では容易とはいえない．

MOJIZOkin は，筆者らが iPhone/Pod touch 用アプリとして構築した画像処理ソフトウェアである．カラー画像から，字形が黒で背景が白となる 2 値画像を生成することを目的としている．小型コンピュータでは，主に画面サイズの制限によって，多数のパラメータを制御する必要のある画像処理は利用が難しい．そこで，明度を用いた字形/背景の分離に加えて，筆者らが古代木簡解読支援のために構築した 1 パラメータで制御可能な画像処理を搭載した．また，複数の画像処理を重畳できるように，各画像処理では字形の一部と推定される画素の色を残し，最後に 2 値化を行う方法を採用した．以上に加えて，タッチ操作による字形/背景の修正，画像の反転もサポートした．ただし，画像のトリミングと回転の機能は iOS 標準の写真アプリで対応できるため搭載していない．MOJIZOkin による処理のフローを図 **10** に示す．

2017 年 5 月 4 日現在，MOJIZOkin は 2,500 を超える Apple ID ユーザによってダウンロードされている．画像処理の選択自体は本論文の本質的論点ではないが，後述の実験における当検索利用時の選択肢の 1 つとしてこのアプリを採用するものとする．

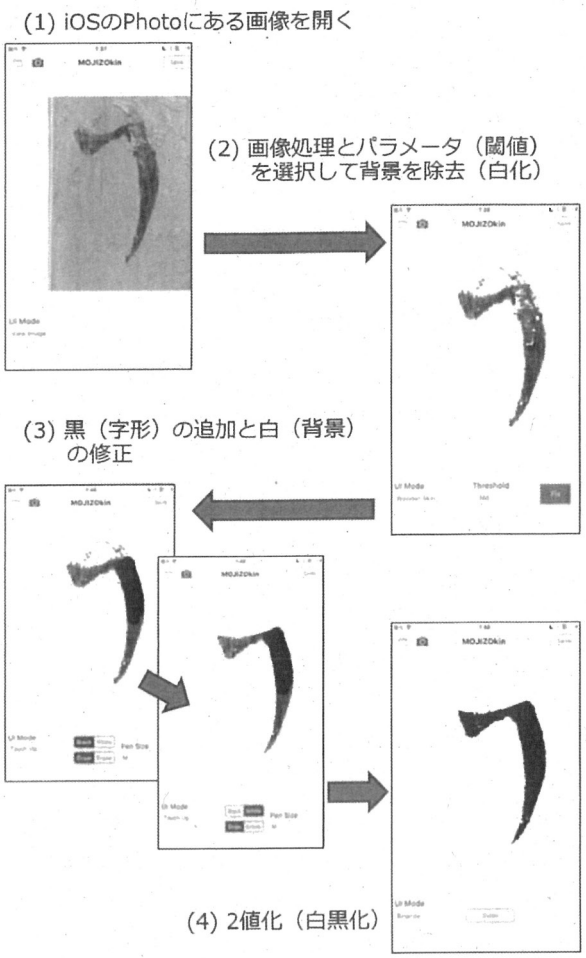

図 10 MOJIZOkin による画像処理
Fig. 10 Image processing using MOJIZOkin.

4. 評価および考察

4.1 MOJIZO によるキー数の推移

古文書デジタルアーカイブの利用状況を示す絶対的な基準の設定は困難である．しかし，MOJIZO については，同じく奈良文化財研究所の Web サイトで公開されている字種をキーとした横断検索（以下，字種検索と記す）との比較を行うことで，相対的ではあるが現実的な評価が可能であると考える．

ここでは，MOJIZO が公開された翌月となる 2016 年 4 月から 12 カ月間に入力されたキーの数を用いて評価を行うものとする．**表 1** に月ごとのキー数を示す．

2016 年 3 月の段階では，新聞などを用いた一般向けのプレスリリースのみを実施している．一方，2016 年 9 月には国際学会での研究者向けの発表 [21] を含めた広範囲への周知を行うとともに，字種検索を含む別ページからのリンクを整備して利便性の改善を図った．このことは，2016 年 10 月以降の MOJIZO のキー数の増加に影響を与えたものと推測している．さらに，2017 年 3 月には画面の小さいクラ

情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

表 1 字種検索と MOJIZO の検索キーの数
Table 1 Numbers of search keys for "字種検索" and MOJIZO.

	字種検索	MOJIZO
2016 年 4 月	21,793	9,453
2016 年 5 月	22,913	7,945
2016 年 6 月	23,470	9,496
2016 年 7 月	22,527	9,348
2016 年 8 月	19,705	8,573
2016 年 9 月	21,665	14,458
2016 年 10 月	23,228	35,081
2016 年 11 月	18,133	37,291
2016 年 12 月	28,430	42,163
2017 年 1 月	25,543	45,973
2017 年 2 月	21,504	41,361
2017 年 3 月	23,942	64,710
合計	272,853	325,852

イアント向けの結果表示方法の変更，および MOJIZOkin のリリースが行われているが，同月のキー数の増加については今後の期間をおいたうえでの検証が必要と考えている．

4.2 古文書の文字の研究者による評価実験

情報検索が妥当な検索結果を出力することは重要な目標であるが，当検索は字種をキーとした検索を補うためのものであり，検索結果の妥当性をキーの字種との一致で評価することは現実的とはいえない．そこで，MOJIZO および MOJIZOkin の構築に関わっていない古文書の文字の研究者 4 名を被験者とする評価実験を行った．このうち 2 名は古代木簡を含む出土文字資料を主に扱っており，別の 2 名は和紙文書を主に扱っている．

本実験に際しては，MOJIZOkin をインストールした第 6 世代の iPod touch（CPU：Apple A8 1.0 GHz，主メモリ 1 GB）を用意した．ただし，被験者がこの機材/アプリを使用するかどうかは自由とした．結果として，被験者 2 名は Windows PC を使用（うち 1 名は iPod touch を併用）した．

被験者は，それぞれ任意でキーを用意し，MOJIZO による検索を 100 回以上実施した．画像処理の利用を含めて，本実験で被験者が検索作業を行った時間はそれぞれ 5 時間程度，あるいはそれ以上であった．そのうえで，MOJIZO の検索結果に対する下記の質問に 5 件法での回答を行った．

検索結果（上位 8 または 6 個）には，検索に使用した画像と「形状」が類似した画像が含まれていましたか
- そう思う
- ややそう思う
- どちらともいえない
- あまりそうは思わない
- まったくそうは思わない

その結果，「そう思う」が 2 名，「ややそう思う」が 2 名となった（複数の機材を併用した被験者の回答は高いものを採用）．ただし，和紙文書を専門に扱う被験者の評価はともに「ややそう思う」であり，「検索する文字画像によって，検索結果にかなりばらつきが出る」，「検索文字自体がヒットしないケースがままある」とのコメントが併記された．

字形検索サーバの精度改善は重要な課題である．また，横断検索ゆえに発生しうる字形画像の特徴差およびテンプレート登録手続きの違いについても検討が必要と考える．和紙文書のデジタル画像は彩度の分布が字形/背景を問わず低く，ノイズと字形を区別した 2 値化の自動化が難しい．また，古代木簡の字形画像をテンプレートとして登録する際には当該文書解読の専門家が 2 値化とノイズ除去を実施しているが，和紙文書の字形画像では隣接文字の字形の混入を含めて専門家によるノイズ除去が実施できておらず，2 値化の結果を専門家が確認できていない字形画像も含まれる．これらは，字種に比べると仕様の共通化が難しい字形画像の横断検索を実装，運用するうえでの課題と認識している．

また，クライアントの画面サイズに応じて Web サーバが検索結果として表示する字形画像数を変更する機能について，先と同じ被験者・機材による評価実験を実施した．iPod touch の液晶パネルは 4 inch，解像度（dot）は 640 × 1,136 で，縦長に使用することで画像入力領域と検索結果となる字形画像を同時に画面内に表示することができた．このとき，字形画像はデジタルアーカイブごとに 6 個で，各字形画像の長辺は 98 dot/約 8 mm であった．字形画像は拡大表示可能だが，その場合は他の字形画像，画像入力領域などを画面外に押し出す必要が生じた．一方，Windows PC は液晶パネルが 24 inch，横 × 縦 の解像度は 1,920 × 1,080，画面の表示倍率は 100%で，画像入力領域と検索結果となる字形画像を同時に画面内に表示することができた．このとき，字形画像はデジタルアーカイブごとに 8 個で，各字形画像の長辺は 100 dot/約 27 mm であった．また，液晶パネルには表示領域に若干の余裕があり，画像入力領域と字形画像を同時表示した状態で 125%での拡大表示が可能だった．ただし，iPod touch，Windows PC のいずれについても，拡大表示に関する被験者への制限，指示は行っていない．被験者は，下記の質問に対して 5 件法での回答を行った．

MOJIZO の検索結果（画像）は木簡字典/電子くずし字字典データベースにリンクしたボタンになっています
検索結果となる画像の数（上位 8 または 6 個）と大きさはいかがでしたか
- 数が多すぎる/画像が小さすぎる
- 数が多い/画像が小さい
- ちょうどよい
- 数が少ない/画像が大きい
- 数が少なすぎる/画像が大きすぎる

情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

その結果，iPod touchだけを使用した被験者のうち1名が「数が多い/画像が小さい」と回答し，残り3名は「ちょうどよい」と回答した．iPod touchの液晶パネルは最新のスマートフォンに比べると小さいが，同等の液晶を備えるスマートフォンの利用者は現時点では多いと推定される．Webサーバの使い勝手に関する改善を通して古文書デジタルアーカイブの活用を進めていくことも課題の1つであると考えている．

5. おわりに

本論文では，字形画像をキーとした情報検索技術が，古文書デジタルアーカイブの利用を促進する効果について述べた．12カ月間の運用において，字形画像をキーとした十分な数の検索が実施されたこと，および字種による検索とは異なるニーズに対応できた可能性が高いことが明らかとなった．現在，筆者らが提供できる環境では，キー数に占める利用者の増加分/利用者あたりの入力回数の増加分の分析は困難であるが，今後の研究活動を通して両方を活性化させるための技術の実現を続けていきたいと考えている．一方で，専門家による評価においては技術面および運用面の課題も明らかとなった．今後の課題として，検索精度と使い勝手の向上，テンプレートとなる字形画像の整理と追加があげられる．テンプレートが増加し，類似した形状のテンプレート群が有効なクラスタを形成できるようになれば，クラスタ内の共通性に着目した検索精度の改善が可能になると考えられる．さらに，各クラスタに識別子を設けることで，字種情報を利用しない字形検索においても，各種の教師あり学習/半教師あり学習の適用，テンプレートマッチング以外の手法による高精度化などへの道が開けると考えられる．

謝辞 評価実験にご協力いただいた研究者の皆様に謹んで感謝の意を表する．本研究は，科学研究費 基盤（S）-25220401，基盤（A）-26244041，基盤（A）-26240049，基盤（C）-15K02841の助成により実施したものである．

参考文献

[1] Core, D.: Metadata Initiative (DCMI), available from ⟨http://dublincore.org/⟩ (accessed 2017-05-05).
[2] Europeana collections, available from ⟨http://www.europeana.eu/portal/en/⟩ (accessed 2017-05-05).
[3] World Digital Library, available from ⟨https://www.wdl.org/en/⟩ (accessed 2017-05-05).
[4] National Digital Archives Program, Taiwan, available from ⟨http://www.ndap.org.tw/index_en.html⟩ (accessed 2017-05-05).
[5] 奈良文化財研究所 木簡字典，入手先 ⟨http://jiten.nabunken.go.jp/⟩（参照 2017-05-05）．
[6] 東京大学史料編纂所：電子くずし字字典データベース，東京大学史料編纂所データベース検索，入手先 ⟨http://wwwap.hi.u-tokyo.ac.jp/ships/db.html⟩（参照 2017-05-05）．
[7] 『木簡画像データベース・木簡字典』『電子くずし字字典データベース』連携検索，入手先 ⟨http://r-jiten.nabunken.go.jp/⟩（参照 2017-05-05）．
[8] Panichkriangkrai, C., Li, L., Walker, R. and Hachimura, K.: Image Analysis for Historical Japanese Book Archives, *International Journal of Asian Business and Information Management*, Vol.5, No.2, pp.1–11 (Apr.-June 2014).
[9] 寺沢憲吾，長崎 健，川嶋稔夫：固有空間法とDTWによる古文書ワードスポッティング，電子情報通信学会論文誌，Vol.J89-D, No.8, pp.1829–1839 (2006).
[10] 文書画像検索システム，入手先 ⟨http://records.c.fun.ac.jp/⟩（参照 2017-05-05）．
[11] 早坂太一，大野 亙，加藤弓枝，山本和明：ディープラーニングによる変体仮名の翻刻およびWWWアプリケーション開発の試み，人文科学とコンピュータシンポジウム論文集，No.2, pp.7–12 (2016).
[12] 北本朝展，山本和明：人文学データのオープン化を開拓する超学際的データプラットフォームの構築，人文科学とコンピュータシンポジウム論文集，No.2, pp.117–124 (2016).
[13] Truyen, P.V., 中川正樹，馬場 基，渡辺晃宏：木簡画像集録システムの設計と実現，日本情報考古学会誌「情報考古学」，Vol.19, No.1·2, pp.1–12 (2013).
[14] 来代誠仁，白井啓一郎，遠藤友樹，中川正樹，馬場 基，渡辺晃宏，井上 聡，久留島典子：古代木簡に対する平滑化処理の適用および古代木簡解読支援システムのアップデート，人文科学とコンピュータシンポジウム論文集，No.4, pp.65–70 (2013).
[15] MOJIZO, available from ⟨http://mojizo.nabunken.go.jp/⟩ (accessed 2017-05-05).
[16] MOJIZOkin, available from ⟨https://itunes.apple.com/jp/app/mojizokin/id1211838518?mt=8⟩ (accessed 2017-05-05).
[17] Tsukumo, J. and Tanaka, H.: Classication of Handprinted Chinese Character Using Nonlinear Normalization and Correlation Methods, *Proc. 9th ICPR*, Roma, Italy, pp.168–171 (Aug. 1988).
[18] Yamada, H., Yamamoto, K. and Saito, T.: A Nonlinear Normalization Method for Handprinted Kanji Character Recognition Line Density Equalization, *Proc. 9th ICPR*, Roma, Italy, pp.172–175 (Aug. 1988).
[19] Liu, C.L., Kim, I.J. and Kim, J.H.: High accuracy handwritten Chinese character recognition by improved feature matching method, *Proc. 4th ICDAR*, Ulm, Germany, pp.1033–1037 (1997).
[20] Liu, C.L.: Handwritten Chinese Character Recognition: Effects of Shape Normalization and Feature Extraction, Lecture Notes in Computer Science, Vol.4768/2008, pp.104–128 (2008).
[21] Kitadai, A., Nakagawa, M., Baba, H. and Watanabe, A.: Similarity Evaluation and Shape Feature Extraction for Character Pattern Retrieval to Support Reading Historical Documents, *Proc. 10th IAPR International Workshop on Document Analysis Systems* (*DAS*), Gold Coast, Australia, pp.359–363 (Mar. 2012).
[22] Kitadai, A., Takata, Y., Inoue, M., Fang, G., Baba, H., Watanabe, A. and Inoue, S.: A Web Based Service to Retrieve Handwritten Character Pattern Images on Japanese Historical Documents, *6th Conf. Japan Association for Digital Humanities* (*JADH 2016*), Tokyo, Japan, Vol.1, p.57 (Sep. 2016). available from ⟨http://conf2016.jadh.org/abstracts/p-12/⟩.

情報処理学会論文誌 Vol.59 No.2 351–359 (Feb. 2018)

耒代 誠仁 （正会員）

2004 年東京農工大学大学院工学研究科博士後期課程修了．同年より同大学研究員，助手，助教，特任准教授，桜美林大学講師を経て，現在，桜美林大学准教授．手書き文字認識技術の応用，コンピュータと教育，古文書解読支援/DB 検索技術等の研究・教育に従事．電子情報通信学会，日本情報考古学会，ヒューマンインタフェース学会各会員．博士（工学）．

高田 祐一 （正会員）

2005 年関西学院大学文学部史学科日本史学専攻卒業．2007 年同大学大学院文学研究科博士前期課程修了．株式会社日本総合研究所等を経て，現在，奈良文化財研究所企画調整部文化財情報研究室研究員．考古学・文献史学におけるデータベース活用および前近代石切場研究に関心がある．修士（歴史学）．

井上 幸

2004 年武庫川女子大学大学院文学研究科博士後期課程単位取得満期退学．奈良文化財研究所都城発掘調査部史料研究室アソシエイトフェロー等を経て，現在，東大阪大学こども学部アジアこども学科准教授．日本古代の字形，日本語史に関心がある．博士（文学）．

方 国花

2012 年愛知県立大学大学院国際文化研究科博士後期課程修了．現在，奈良文化財研究所都城発掘調査部史料研究室アソシエイトフェロー．古代東アジアの出土文字資料に使われる漢字字体に関心がある．博士（日本文化）．

馬場 基

1995 年東京大学文学部卒業．2000 年同大学大学院人文社会系研究科博士課程中退．現在，奈良文化財研究所都城発掘調査部主任研究員．平城宮・京跡の発掘調査や出土文字資料の整理・調査・研究，情報発信に従事．専門は，日本古代史・木簡学等．修士（文学）．

渡辺 晃宏

1982 年東京大学文学部国史学科卒業．1989 年同大学大学院博士課程単位取得退学．現在，奈良文化財研究所副所長・都城発掘調査部副部長・史料研究室長．平城宮・京の発掘調査と出土文字資料の研究に従事．木簡学会会員．文学修士．

井上 聡

1992 年東京大学文学部国史学科卒業．1998 年同大学大学院人文社会系研究科博士課程単位取得退学，現在，東京大学史料編纂所助教．日本中世史専攻．中世古記録の編纂を主務としつつ，データベースの構築にも従事．研究は荘園史・社会経済史を主対象とする．修士（文学）．