訓点資料における翻刻支援システムの構築

田中 勝（和歌山大学大学院 システム工学研究科）
村川 猛彦（和歌山大学 システム工学部）
宇都宮 啓吾（大阪大谷大学 文学部）

訓点資料の解読支援環境を確立するために開発を行ってきた，デジタルアーカイブシステムについて報告する．対象とする資料は手書き文書であり，書写後に漢文訓読のためヲコト点などの訓点が書き加えられている．そのような資料画像を翻刻し，資料とテキストとの対応づけを行う際には，大きさが不揃いな漢字・送り仮名・ヲコト点が混在している点を考慮する必要がある．構築した支援システムおよびデータベースでは，文字および訓点を，画像における文字領域や点座標として処理しており，これによりブラウザを用いて，画像上の操作が可能な入力インターフェイスを実現した．

Transcription Support System for Written Materials with Reading Marks

Masaru Tanaka (Graduate School of Systems Engineering, Wakayama University)
Takehiko Murakawa (Faculty of Systems Engineering, Wakayama University)
Keigo Utsunomiya (Faculty of Literature, Osaka Ohtani University)

A digital archive system developed for deciphering written materials with reading marks is described. Intended documents include guiding marks such as *okototen* attached after the body texts were handwritten. When computerizing those materials to associate the content with the text data, we require consideration of the fact that variably-sized characters and symbols are mixed on a document. Since the support system and the database that we have constructed treat the formation elements by means of coordinates over the material image, we implemented an input interface so that the users can operate on the image via their browsers.

１．はじめに

国語学や仏教学において，奈良・平安・鎌倉時代の言語や文化を知るために，訓点資料の解読が進められており，解読作業の効率化・高品質化を支援する環境が望まれている[1]．現在，様々な古写経の電子化が進められている中で，訓点資料については電子化がほとんど進んでいないのが現状である．そのため，解読作業の電子的な支援や学生などの教育を支援する環境の整備が遅れており，この点については，例えば，訓点資料を研究対象とする訓点語学会において，今年度初めて訓点資料講習会が開催される（2015 年 9 月，東京大学）など，当該分野における後継者育成や環境整備に対する意識が高まっているところであり，訓点資料の電子画像化および電子テキスト化とともに，それを扱うシステムの提供が求められている．

そこで，筆者らは訓点資料のデジタルアーカイブ化を通じて，解読支援環境の確立を目指し，デジタルアーカイブシステムの開発を行ってきた．まさに，時期に即した試みとして位置づけられるものと考えている．本稿では，これまで行ってきた訓点資料のテキスト化を支援する機能の実装や，それを扱うデータベースシステムの構築について報告する．

２．訓点資料のデジタルアーカイブ

２．１ 訓点資料とは

訓点とは，漢文の訓読のために書き加えられた仮名や諸符号（句読点，返り点，ヲコト点）をいい，訓点資料とは，それらの訓点を付けられた資料のことである．諸符号のうち，ヲコト点とは，頻出する助詞や助動詞の代用として用いられた表記で，その働きは，漢字に対する打点位置や点の形によって決まる．訓点資料は，奈良時代の写経を始めとし，平安，鎌倉時代へと漢文に訓点を書き加える文化が続いている．中でも平安時代が顕著であり，この時代の資料だけでも 5,000 点以上現存していると言われている．そういった歴史の中で定められてきたヲコト点の形式は，様々な宗派・流派等によってそれぞれ異なっており，それらを特定することで，資料の来歴などを突き止めることができる．

訓点資料の一例として，『千手千眼陀羅尼経残巻』を挙げる．図 1 は，この経典画像の一部と文字の詳細である．この経典は，奈良時代の古写経であり，国宝として京都国立博物館に展示され，画像が e 国宝にて Web 公開されている[2]．資料の文書には，漢字・送り仮名・ヲコト点が見られる．なお，『千手千眼陀羅尼経残巻』で用いられているヲコト点の形式は宝幢院点で，約 40 種類の形状と打点位置の組み合わせにより，多数の読みが存在する[3]．頻出する形状における打点位置と読みの例を図 2 に示す．

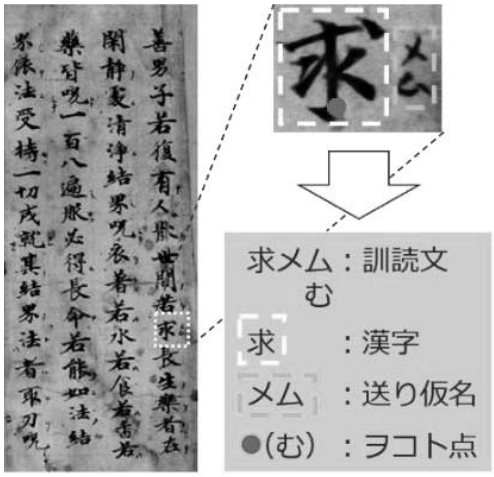


図 1 経典画像の一部と文字の詳細

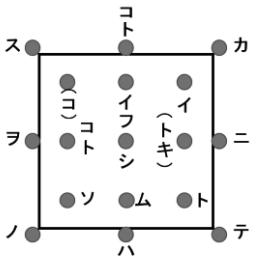


図 2 宝幢院点における点の読みの一例

２．２ アーカイブ化にあたっての課題とその解決案

訓点資料を読解する際の課題として，大きさが不揃いな漢字・送り仮名・ヲコト点が混在していること，訓点資料のテキスト化が十分になされていないこと，学生など専門家以外の作業への参加が難しいことが挙げられる[4]．

また，テキスト化における課題として，文書情報が雑多であるため一つ一つデータ入力することがかなりの手間となること，ヲコト点の打点位置が曖昧で機械的に読みを確定しにくいこと，墨を使った手書きに訓点が加わった資料なので計算機による文字認識は容易ではないことが挙げられる．

テキスト化の課題解決にあたって，本研究では利用者と計算機で役割分担を行うことで，効率の良いテキスト化を目指す．利用者は，マウス操作により漢字・送り仮名・ヲコト点の位置や領域を入力し，計算機は，それに基づいて各文字領域の行・列番号を算出するほか，漢字と訓点，ヲコト点と読みの関連付けを行う．利用者が読み（打点位置）を選択することもできる．

３．システム構築

本研究では，訓点資料の解読支援環境の構築を目指し，デジタルアーカイブシステムの開発・提供を行う．そのためには，訓点資料のテキスト化を支援する機能の実装や，それを扱うデータベースシステムの構築が不可欠である．本稿では，これまで開発を行ってきたテキスト化を支援するためのシステム[5]について詳しく述べるとともに，新たに作成した，ヲコト点のテキスト化支援機能について報告する．

３．１ システム構成

図 3 にシステム構成図を示す．Web アプリケーションは HTML5（Canvas），JavaScript，jQuery を用いており，jQuery のライブラリファイルの他，4 ファイル約 2,000 行の自作プログラムでインターフェイスを構築した．特徴としては，画面遷移なしで文書情報の取得と表示を行うことができるほか，クロスブラウザに対応していることが挙げられる．

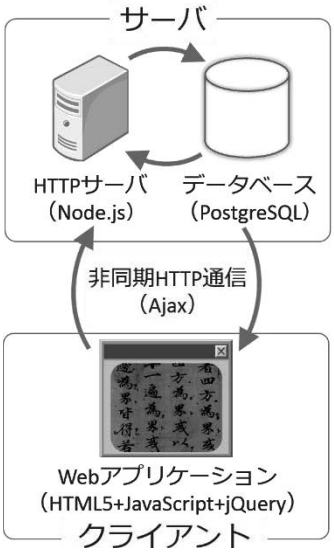


図 3 システムの構成図

HTTP サーバは Node.js を用いており，文書情報の保存と読み出しを行う．クライアント・サーバ間の通信には Ajax を使用することで，非同期通信によりスムーズな操作が可能となっている．

データベースには PostgreSQL を導入しており，訓点資料に対応したデータベースを設計し，データを管理するようにした．

３．２ データベース設計およびデータ形式

データベースの設計について説明する．訓点資料の特徴として，本文中の漢字に訓点（送り仮名やヲコト点など）が付与されていることが挙げられる．この特徴をデータベースに対応するために，取得する文書情報に対して 1 文字・1 箇所単位で ID を設定し，漢字と訓点のそれぞれの関係を ID で紐付けする．こうすることで，どの訓点はどの漢字に属するものかがわかり，1 つの漢字に対して複数の訓点が付与されている場合でも対応することができる．

扱うデータの形式について説明する．文書情報は，画像上の座標をベースに，文字領域や点座標としてデータベースに保存している．座標で情報を扱うことで，曖昧な手書き文書を計算機上で扱えるようにした．また，座標で文書情報を扱う上でのメリットとして，得られた座標から漢字と訓点の位置関係を推定し自動で文書情報を得られることや，一領域ごとの画像を容易に取得できること，などが挙げられる．前者は，後述の自動関連付け機能としてシステムに反映しており，後者は，Canvas や jQuery の機能を利用して，一箇所単位での画像保存を行っている．なお，得られた画像は PostgreSQL で扱うために，Base64 を用いて文字列に変換している．変換された文字列は，「（略）TkSuQmCC」のような記述となっており，幅 144 ピクセル，高さ 120 ピクセルの PNG (Portable Network Graphics)形式で，約 45KB になる．

座標処理の応用例として，得られた漢字の領域情報を元に，『千手千眼陀羅尼経残巻』の一部の漢字に対してトリミングを行った例を図 4 で示す．

図 4 領域情報を用いたトリミング例

ヲコト点の読み情報は，形状・打点位置ごとに分けて保持しておくことで，図 2 のような読み情報をデータベースに反映させることができる．ヲコト点の形状は，現代の漢字・カタカナ文字や記号をそのまま表示するか，それらを回転・反転させて表示することで表現することができる．データベース上では，文字・記号と回転・反転の有無の組み合わせで，ヲコト点の形状を記録している．

３．３ システムの利用想定

システムの動作環境について，今回は複数のクライアントによる並列作業は考慮せずに，作業者が 1 人であることを想定してシステムを構築した．また，サーバとクライアントは同一の LAN（Local Area Network）内にあることを想定している．システムのレスポンスについては，クライアントの起動時に資料画像を読み込むために数秒の時間がかかるものの，操作時には特に問題なくスムーズに入力作業が行うことができる．ただし，最終的に扱う資料の規模として，一文書数千字の入力を想定しているが，負荷テストは現時点では行っていないため，レスポンスの変化を確認する必要がある．

３．４ システムの入力インターフェイス

これまで，ブラウザベースの入力インターフェイスの開発を行ってきた．図 5 にブラウザ画面例を示す．インターフェイスの基本機能として，マウスを使った入力機能や，ヲコト点のテキスト化支援機能，画面遷移を伴わない画像移動機能などが挙げられる．

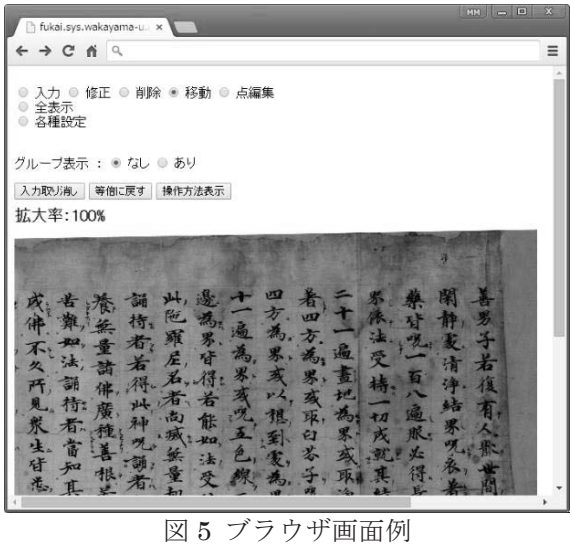
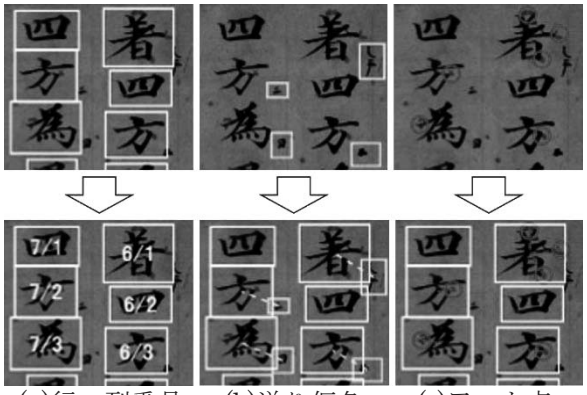


図 5 ブラウザ画面例

入力機能について説明する．ブラウザ画面の資料画像上で，文字の領域をドラッグしたり，ヲコト点の位置をクリックしたりするといったマウス操作を通じて，位置・領域情報を簡単に入力す

ることができる．また，関連付け機能により，漢字の行・列番号や漢字と訓点との関連といった文書情報は，入力で得られた座標を元に自動で取得される．図 6 は，入力例（図上部）と関連付け機能の例（図下部）である．行・列番号は「列番号／行番号」の形式で表示され，漢字と訓点の関連情報はそれぞれを点線で結び表示している．



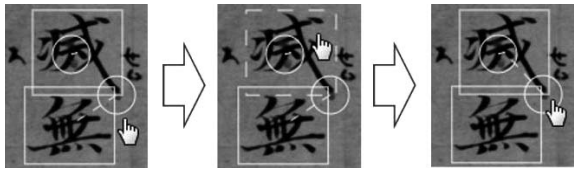
(a)行・列番号　(b)送り仮名　(c)ヲコト点

図 6 入力例と自動関連付け機能例

データの削除・修正においても，マウス操作で簡単かつスムーズに行うことができる．削除に関しては，指定削除機能を使用し，対象となる漢字または訓点の場所をクリックするだけで削除を行うことができる．図 7 がその例である．次に，漢字と訓点の関連情報の修正は，関連情報の指定修正機能を使用し，正しい漢字と訓点の組をそれぞれクリックすることで，直ちに修正することができる．図 8 の例では，漢字「滅」の右下にあるヲコト点「・」が漢字「無」に誤って関連付けられている．この例では，まず「滅」の領域をクリックした後に，そのヲコト点「・」のマーカーをクリックすることで，関連付けを修正している．

(a)削除前　(b)削除指定と確認

図 7 入力データの指定削除機能



(a)修正前　(b)修正中　(c)修正後

図 8 関連情報の指定修正機能例

入力以外のマウスを使った操作として，画像上でマウスホイールを使うことで，図 9 のように画像の拡大・縮小ができる．また，画像移動機能により，画面上部のメニューで移動モードを選択し，マウスを画像上でドラッグすることで，経典画像を自在に移動させることができる．図 10 は，画像を左下から右上方向に移動させた例である．また，図 11 に示した設定画面から，資料画像上の表示（データの表示色，ヲコト点のマーカーサイズ，関連付け情報の表示有無など）を変えることで，資料に合わせた表示や利用者の使用感に合わせた表示設定を行うことができる．

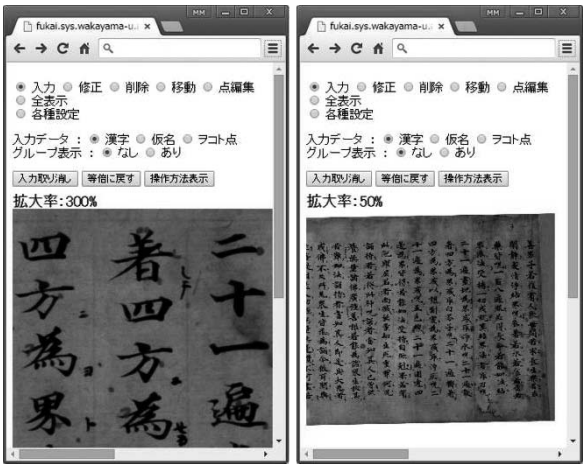


図 9 資料画像の拡大縮小機能例

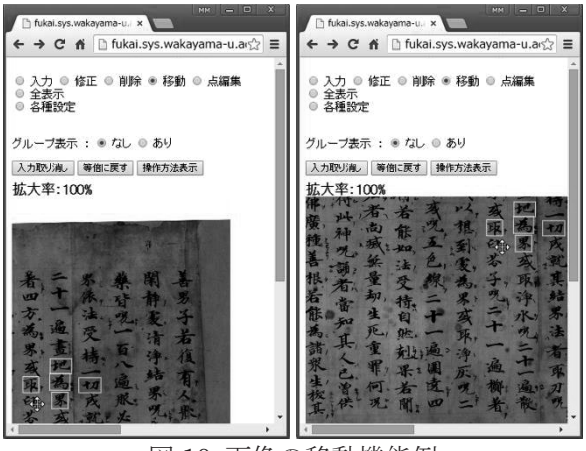


図 10 画像の移動機能例

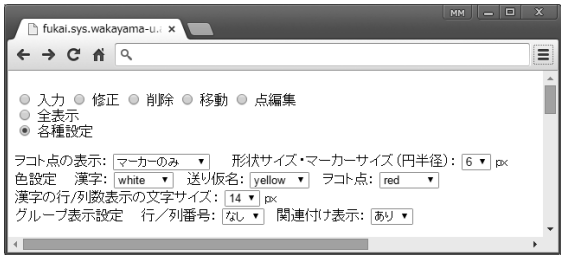


図 11 データの表示設定

3．5 ヲコト点の読み入力機能と操作手順

ヲコト点のテキスト化支援機能は，マウスを使った入力機能，ヲコト点を見やすくするための文字拡大機能，形式・形状ごとの読みと打点位置の対応表示機能などで構成される．

操作の流れについて説明する．まず，画像上でヲコト点の読み入力対象の文字の領域をクリックすると，対象文字の拡大画像が表示される（図 12．なお，図中の吹き出しは説明のため付けたもので，実際の画面には表示されない．以下同じ）．次に，文字画像上のヲコト点をクリックすると，ヲコト点の形状に対応した読みと打点位置の表が表示される（図 13）．そして，表に表示された読みをクリックし選択することで，ヲコト点の読みが登録される（図 14）．登録された読みは，漢字の拡大画像やヲコト点の形状と共に確認することができる（図 15）．

図 12 ヲコト点の読み入力例(1)

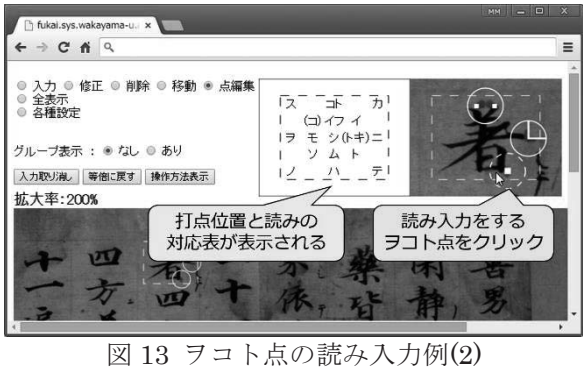


図 13 ヲコト点の読み入力例(2)

図 14 ヲコト点の読み入力例(3)

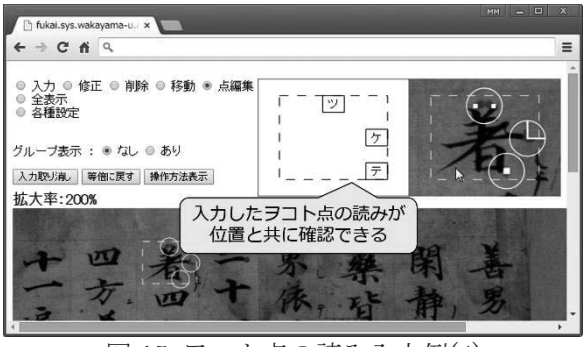


図 15 ヲコト点の読み入力例(4)

4．関連研究

訓点資料のテキスト化を支援するシステムの先行研究には，田島ら[6]がある．この事例では，テキストベースのインターフェイスを採用しており，訓点情報は本文テキストに付与することで，資料のテキスト化を行っている．研究課題として，インターフェイスのテキスト上の打点位置と資料上の読みとの関係がずれてしまうことが指摘されている．

それに対して，本研究では，画像ベースのインターフェイスを採用することとし，資料画像上の位置・領域に文字情報を付与することで，資料画像を見ながらの直感的な操作を可能となっている．前出の課題に対しては，読みの確定は打点位置から機械的に行うのではなく，人が資料画像を見ながら判断するものとし，入力作業に合わせて形式・形状ごとの読みと打点位置の対応表示をリアルタイムに行うことで判断をサポートする．

ヲコト点の打点位置と読みの入力形式に関して，両者を比較する．前者の入力インターフェイスには，入力対象のテキストとテキスト上にグリッドが表示される．利用者は，資料画像のヲコト点を見て，それと同じ位置のテキスト上のグリッドを選択し，打点位置を入力する．それにより，ヲコト点の読みは，入力された打点位置から機械的に確定させる．この方式のメリットは，入力の効率が良く利用者の負担を軽減できることである．デメリットは，読みと打点位置の対応づけを

機械的に行っているために，それらにズレが生じる場合があることである．それに対して後者（本システム）では，入力インターフェイスには，入力対象の文字画像のほか，ヲコト点の打点位置と読みの対応表が表示される．利用者は，資料画像や文字画像を見て，適切な打点位置，または読みを選択し入力する．この方式のメリットは，打点位置，または読みは人が判断するので，手書き文書の曖昧さを解消できることにある．デメリットは，打点位置や読みの判断を一つ一つの入力に対して行わなければいけないため，利用者に入力の負担がかかることである．

５．おわりに

訓点資料の解読支援環境の構築・提供のために，テキスト化を支援するシステムの開発とそれを扱うデータベースシステムの構築を行ってきた．

利用者が最低限の入力を手動で行い，計算機がそれに基づいて文書情報を自動で取得する仕組みを取り入れることで，入力作業の省力化や，手書き文字，ヲコト点に対する柔軟な対応を図った．また，漢字・訓点の文書情報を，文字領域・点座標で処理することで，手書き文書である訓点資料を計算機上で扱うことができた．データベースの設計にあたっては，漢字・訓点をそれぞれ ID 管理することで，訓点資料の情報を文字・訓点の単位で扱えるようにした．ヲコト点のテキスト化については，文字の拡大により資料を見やすくするとともに，読みと打点位置の対応表示をさせることで入力支援を図った．

ヲコト点の入力機能に関して，現時点では，ヲコト点の色入力機能の実装や，データベース上のヲコト点の読み情報編集・修正機能の実装ができていないが，今後より多くの訓点資料を扱えるようにするために，機能実装を進めていく予定である．

今後は，ユーザに対する有益なデータの提供方法の検討，『千手千眼陀羅尼経残巻』以外の訓点資料への適用，利用者を交えたシステムの評価実験を予定している．

参考文献

1）林寺正俊：日本古写経データベースの構築とその意義，情報処理学会人文科学とコンピュータシンポジウム論文集，Vol.2012，No.7，pp.11-16（2012）.
2） e 国宝『千手千眼陀羅尼経残巻』．http://www.emuseum.jp/detail/101016/000/000（参照 2015-09-17）.
3） 築島裕：平安時代訓點本論考 ヲコト點図仮名字体表，汲古書院（1986）.
4） 小助川貞次：漢文訓読史概説の構想，富山大学人文学部紀要，No.56，pp.109-121（2012）.
5） 田中勝，村川猛彦，宇都宮啓吾：訓点資料を対象としたデジタルアーカイブシステムの構築，2015 年電子情報通信学会総合大会，情報・システム講演論文集 1，p.44，D-4-13（2015）.
6） 田島孝治，堤智昭，高田智和：ヲコト点電子化のためのデータ構造と入力支援システムの試作，情報処理学会人文科学とコンピュータシンポジウム論文集，Vol.2012，No.7，pp.211-216（2012）.