

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 9 日現在

機関番号：32605

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24520771

研究課題名(和文)人と英知をユビキタスに結ぶクラウド型古文書デジタルアーカイブのための情報検索技術

研究課題名(英文) Information retrieval technologies of cloud historical document databases for ubiquitous connection between human and wisdom.

## 研究代表者

末代 誠仁 (Kitadai, Akihito)

桜美林大学・総合科学系・講師

研究者番号：00401456

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：本研究で達成した成果は次の通りである。(1) 奈良文化財研究所の「木簡字典」と東京大学史料編纂所の「電子くずし字辞典」を一つの検索キー(字形画像)で検索可能な横断的の字形検索システムを実現した。(2) 画面サイズが数インチの携帯タブレットから、20インチクラスの大画面タッチパネルまで対応するスケーラブルな字形検索システムを提案、実現した。(3) 様々なユーザ/クライアント環境とデジタルアーカイブを有機的に結合する技術的基盤を確立することができた。

研究成果の概要(英文)：This research presents the following results. (1) Providing cross information retrieval technologies for Japanese major digital archives of historical documents: "Mokkan jiten" by Nara national research institute for cultural properties and "Denshi kuzushi-ji jiten" by historiographical institute the university of Tokyo. (2) Proposing scalable user interfaces of character pattern retrieval for variable sizes of touch display. (3) Connecting users/client systems and the digital archives for historical documents coordinately.

研究分野：情報考古学

キーワード：日本史 歴史学 デジタルアーカイブ 情報検索 パターン認識 ユーザインタフェース

### 1. 研究開始当初の背景

古文書は過去を記録した史料であると同時に、先人の英知を現代、そして未来に伝える生きた書物である。古文書デジタルアーカイブの収録点数が飛躍的に増加しつつある今、古文書に整理・記録された先人の英知を時代/文書の種類/用途を超えて活用する情報検索技術の高次化は、古文書の真の価値を埋没させることなく現代・未来に伝える上で不可欠である。

研究代表者は、奈良文化財研究所の史学者らと共に古代木簡デジタルアーカイブの一つ「木簡字典」の構築を進めてきた。8世紀前後を中心に作成・使用された古代木簡は平城宮跡周辺だけで20万点超、日本全体では約35万点が発見されており、当時の言語/地名/人名/産業/地域間交流などを伝える極めて有用な参考史料となっている。木簡字典ではこれらの古代木簡から得られた様々な情報をWeb上などで広く公開している。

また、木簡字典の活用を促進する取組みとして、難読木簡解読支援のための検索技術の研究を進めてきた。汚損・破損による情報欠落が著しい古代木簡を解読する上で古文書デジタルアーカイブは有用な類例の宝庫である。研究代表者らはキー文字列の誤字/脱字/語順変動に柔軟に対応できる曖昧テキスト検索のためのExtended Aho-Corasick法、不完全な字体をキーとする破損字体検索のためのテンプレート修正法[3]などを実現するとともに、これらの技術を搭載した統合解読支援環境「Mokkanshop」の一般公開を通して、極限状態にある古代木簡に残る僅かな情報(断片的な釈文/不完全な字体など)から必要な古代木簡に辿り着ける知的情報検索の実現が可能であることを証明してきた。

これらの研究は古代木簡という領域に特化したものであった。それらが一定の成果を上げつつある中で、文字が持つ普遍性に着目し、さらに広範囲の情報を容易かつ効果的に検索する技術の構築は必然である。このような考え方に基づき、研究代表者らはこれまでに木簡字典と東京大学史料編纂所が管理する「電子くずし字字典データベース」との連携テキスト検索を実現し、時代/文書の種類を超えた古文書デジタルアーカイブの一例を示した。また、紙/布など木片以外の古文書にも広く見られる字形輪郭部の滲み・擦れに頑健な字形解析技術、古文書の腐食、変質などに伴う汚損・変色を字形と高精度に分離する高度画像処理技術などを実現し、知的情報検索の適用範囲拡大と高精度化に向けた高い可能性を示した。

### 2. 研究の目的

研究代表者は、本研究の遂行を通して、(a)破損字体検索/文脈処理/高度画像処理によるデジタルアーカイブの動的インデクシングが時代/文書の種類を超えて有用であること、(b)利用環境に応じたスケーラブルなハイブ

リッド情報検索インタフェースが古文書デジタルアーカイブ群の横断的活用を促進すること、そして(c)高次情報検索技術の実現と提供が多様な利用環境/ユーザーニーズに応えるクラウド型古文書デジタルアーカイブ構築の鍵となることを明らかにすることの3点を目標として設定した。それぞれ、細目については以下に補足する。

(1)高度画像処理と破損字体検索は検索キーとデジタルアーカイブの各レコードをアノテーションによって対話的に結合する。また文脈処理は検索文字列とメタデータの柔軟な結合を提供する。動的インデクシングといえるこれらの特性は時代/文書の種類に依存せず、既存のメタデータによる静的インデクシングと融合することで有用な情報検索が実現できる。ただし、実装レベルでは各デジタルアーカイブの長所を活かすカスタマイズが必要となる。これらを切り分け、諸課題を明確にしながら、実際に複数のデジタルアーカイブを対象とした高次情報検索技術を構築し、それらの有用性を高めることができることを証明する。

(2)大画面ワークステーションから小型携帯端末まで様々な機材を結ぶ高速ネットワークの普及により古文書デジタルアーカイブの利用環境は多様化している。これを古文書デジタルアーカイブの利用スタイルに多様性を生む好機と捉え、電子ペンによるユーザインタフェース構築/ネットワークを利用した協調的古文書解読作業支援の研究成果と経験を活かしつつ、Java VM/Servletなどを基盤としたセキュアで柔軟性に富む負荷分散/機能選択型の情報検索インタフェースを利用者に幅広く提供することで、古文書デジタルアーカイブの持つ書物的価値の向上と利用者間の相互交流の促進が図れることを明らかにする。

(3)古文書デジタルアーカイブは遺物の保管庫に留まらず、幅広い時代の古文書解読支援、文化の考証などに利用可能な参考史料の宝庫である。本研究の遂行を通して多様な情報検索の経路を確保することにより、デジタルアーカイブ群と多数の利用者を有機的に結合する新時代のクラウド型古文書デジタルアーカイブが実現できることを証明する。

### 3. 研究の方法

研究代表者はデジタルアーカイブ管理者である奈良文化財研究所(渡辺晃宏氏、馬場基氏ら)、東京大学史料編纂所(井上聡氏、久留島典子氏)、検索技術に不可欠な画像処理などの研究者である信州大学白井啓一郎氏、パターン認識/ユーザインタフェースの

専門家である東京農工大学中川正樹氏らとの研究協力を行い、効果的な研究目標の達成を実現してきた。また、研究代表者らがこれまでの研究で得たアノテーション技術、画像処理技術、パターン認識技術などを改良、発展し、着実な成果構築にも努めてきた。

#### 4. 研究成果

本研究では、前述の「研究の目的」で挙げた各項目に対して、次のような成果を得ることができた。

- (1) 奈良文化財研究所の「木簡字典」と東京大学史料編纂所の「電子くずし辞典」を一つの検索キー（字形画像）で検索可能な横断的字形検索システムを実現した。クライアント・サーバ構成によるこのシステムでは、クライアント側を担当するユーザインタフェース、各デジタルアーカイブを担当する字形検索サーバ、およびこれらを結合するネットワーク/プロトコルで構成される。クライアント、サーバを独立して開発、更新することが可能であり、字形情報の更新、最新の画像処理・ユーザインタフェースの迅速な導入を実現しつつ、複数のデジタルアーカイブへの動的インデクシングの提供を行うことができた。
- (2) 画面サイズが数インチの携帯タブレットから、20インチクラスの大画面タッチパネルまで対応するスケーラブルな字形検索システムを提案、実現した。アノテーションを含む各種操作はマウスだけでなくタッチでも可能なものとした。また、特に処理速度・主記憶領域の制限が大きい携帯タブレットに対してはクライアント・サーバ方式による検索処理支援を行うことで実用性にも配慮した。
- (3) 上記の(a)および(b)に対する成果を組み合わせることで、様々なユーザ/クライアント環境とデジタルアーカイブを有機的に結合する技術的基盤を確立することができた。

これらの成果については、査読付を含む国内外の論文誌・研究会などでの公開・発表を通して広範囲な普及にも尽力した。

#### 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, 白井啓一郎, 馬場基, 渡辺晃宏, 井上聡, 久留島典子, 中川正樹, 古文書デジタルアーカイブに対する横断的字形検索サービスの試作, *情報処理学会 人文科学とコンピュータシンポジウム「じんもんこん 2014」論文集*, Vol.2014, No.3, pp.87-92, (2014), 査読有,

[https://ipsj.ixsq.nii.ac.jp/ej/?action=pages\\_view\\_main&active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=107391&item\\_no=1&page\\_id=13&block\\_id=8](https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=107391&item_no=1&page_id=13&block_id=8).

- [2] 未代誠仁, 白井啓一郎, 馬場基, 渡辺晃宏, 井上聡, 久留島典子, 中川正樹, 古文書字形検索サーバの設計と試作, *日本情報考古学会講演論文集*, Vol.2014, No.13, pp.75-77, (2014), 査読有.
- [3] 未代誠仁, 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, *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 2013」論文集*, Vol.2013, No.4, pp.65-70, (2013), 査読有, [https://ipsj.ixsq.nii.ac.jp/ej/?action=pages\\_view\\_main&active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=96420&item\\_no=1&page\\_id=13&block\\_id=8](https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=96420&item_no=1&page_id=13&block_id=8).
- [4] K. Shirai, Y. Endo, A. Kitadai, S. Inoue, N. Kurushima, H. Baba, A. Watanabe, M. Nakagawa, Character Shape Restoration of Binarized Historical Documents by Smoothing via Geodesic Morphology, *Proc. 12th International Conference on Document Analysis and Recognition*, Vol.1, pp.1317-1321, (2013), reviewed.
- [5] Akihito Kitadai, Keiichi Shirai, Satoshi Inoue, Noriko Kurushima, Hajime Baba, Akihiro Watanabe and Masaki Nakagawa, Image processing and shape repairing for historical character pattern retrieval, *Proc. 16th Biennial Conference of the International Graphonomics Society*, Vol.1, pp58-61, (2013), reviewed.
- [6] 白井啓一郎, 未代誠仁, 井上聡, 久留島典子, 馬場基, 渡辺晃宏, 中川正樹, 古文書字形検索のための画像処理, *情報処理学会「人文科学とコンピュータ研究会(CH-97)研究報告」*, Vol.2013-CH97, No. 7, pp.1-6, (2013), 査読無.
- [7] 未代誠仁, 白井啓一郎, 井上聡, 久留島典子, 馬場基, 渡辺晃宏, 中川正樹, シームレスコンピューティングのための古文書字形検索技術, *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2012)論文集」*, Vol.2012(7), pp85-92, (2012), 査読有, (情報処理学会山下記念研究賞受賞), [https://ipsj.ixsq.nii.ac.jp/ej/?action=pages\\_view\\_main&active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=87022&item\\_no=1&page\\_id=13&block\\_id=8](https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=87022&item_no=1&page_id=13&block_id=8).

〔学会発表〕(計 6 件)

- [1] 末代 誠仁, 古文書デジタルアーカイブに対する横断的字形検索サービスの試作, 情報処理学会「人文科学とコンピュータシンポジウム(じんもんこん 2014)」, 2014 年 12 月 13 日, 国立情報学研究所(東京都).
- [2] 末代 誠仁, 古文書字形検索サーバの設計と試作, 日本情報考古学会 第 33 回大会, 2014 年 9 月 27 日, 東京農工大学(東京都).
- [3] 末代 誠仁, 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, 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 2013」, 2013 年 12 月 12 日, 京都大学(京都市).
- [4] 末代 誠仁, 木簡研究の過去・現在・未来 質疑応答・自由討論など, 歴史の証人 木簡を極める(後援, 文化庁ほか), 2013 年 9 月 22 日, 有楽町朝日ホール(東京都).
- [5] Akihito Kitadai, Image processing and shape repairing for historical character pattern retrieval, *16th Biennial Conference of the International Graphonomics Society*, June-11-2013, Todaiji (Nara, Japan).
- [6] 末代 誠仁, シームレスコンピューティングのための古文書字形検索技術, 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 2012)」, 2012 年 11 月 17 日, 北海道大学(札幌市).

〔その他〕

桜美林大学教員 Web ページ:

<http://gproweb1.obirin.ac.jp/obuhp/KgApp?kyoinId=ymdigiyyiggy>

## 6. 研究組織

### (1) 研究代表者

末代 誠仁 (Kitadai, Akihito)

桜美林大学・総合科学系・講師

研究者番号: 00401456