

木簡など出土文字資料釈読支援システムの高次化と 総合的研究拠点データベースの構築

The improvement of the support system to decode
excavated historical documents and the development of
the comprehensive research center database

渡辺 晃宏 (WATANABE AKIHIRO)

独立行政法人国立文化財機構奈良文化財研究所・都城発掘調査部
・史料研究室長



研究の概要

推論機能を有する木簡の文字認識システム「Mokkanshop」と木簡の文字画像データベース「木簡字典」を中核とし、多様な知識データベースを有機的に結合させた出土文字資料研究拠点データベースを構築する。そして、木簡情報の知の蓄積が解読を可能にし、それが新しい知としてさらなる解読を生み出すという、「知のスパイラル(循環)」を実現する。

研究分野：人文学

科研費の分科・細目：史学・日本史

キーワード：古代史、木簡、出土文字資料、データベース、漢字、文字認識、OCR

1. 研究開始当初の背景

奈良文化財研究所において40年以上にわたって培ってきた経験と勘が頼りの木簡調査のノウハウを、普遍的なものとして形に残し広く社会に還元すると同時に、調査の効率化を図るため、基盤研究(S)「推論機能を有する木簡など出土文字資料の文字自動認識システムの開発」(H15～19年度)により木簡の文字認識システム「Mokkanshop」と文字画像データベース「木簡字典」を開発した。

2. 研究の目的

既開発の「Mokkanshop」と、木簡の文字画像データベース「木簡字典」を踏まえ、両者を単独で機能させるだけでなく、これらを中心に有機的な連関をもつ多様な知識データベース群を配置し、木簡など出土文字資料に関する研究拠点データベースを開発するものである。知の効率的な蓄積・利用を図ることにより新しい知を実現するという「知のスパイラル」を実現し、出土文字資料研究のための拠点機能の構築をめざした。

3. 研究の方法

「Mokkanshop」を基礎に認識精度の飛躍的向上と汎用性の構築を図り、複数地点での同時運用も可能な総合的釈読支援システムを構築する。また「木簡字典」のデータの充実に加え、多様な知識データベースを開発して相互連携を図り、「木簡字典」と「Mokkanshop」を中核とする出土文字資料研究拠点データベースを構築する。さらに、他機関のデータベースとの連携を進め、利便性向上を図る。

4. これまでの成果

Mokkanshop の高次化 2010年3月、木簡の文字自動認識システム「Mokkanshop」を商標登録した(第5307067号)。使用するデータ(字書)の改善の他、欠損部分の情報を利用者が付加するシステムを開発し、利用者の思考の柔軟な反映が可能な情報補完技術が実現し、認識率が向上した。これを受け、2010年10月にMokkanshop新版を公開した。

また、Mokkanshop から文字画像データベース「木簡字典」の参照を可能にし、研究拠点データベースとしての連携を整えた。木簡画像の視認性を向上させるために、墨痕と背景を分離する画像処理技術の研究を行い、その成果は、国際学会ICFHR2010において、ITESOFT best poster paper awardを受賞した。

墨書土器・漆紙文書など木簡以外の出土文字資料の文字認識や、正倉院文書や奈良朝写経の文字の字書として活用も検討中である。

釈読支援システムの同時運用方式の開発 Microsoft SharedViewを用い、東京大学史料編纂所との間で赤外線観察装置の画像共有実験を行った。若干の時間差が発生するものの、問題なく使用できることがわかったが、音声共有については課題を残した。

アノテーションツールの開発 削屑の効率的整理・観察用に、削屑のデジカメ画像に管理情報や釈読成果を書き込めるアノテーションツールを開発し実用化の目途を得た。

文化財撮影用レンズの開発 赤外線と可視光の撮影が同一工程内のフィルタ切り替

えだけで可能な特殊仕様レンズを開発し、木簡の画像記録を大幅に効率化した。

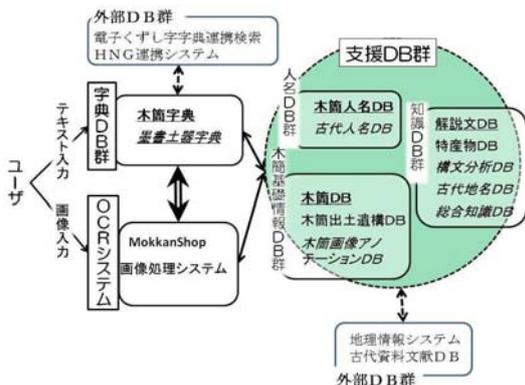
文字画像データの蓄積とXMLの導入 木簡の文字画像データベース「木簡字典」の充実のため、文字画像切り出しシステム Mojiga を改良し、時代・地位を越えたデータの蓄積に努めた。2010年度までに約28,000点の画像を蓄積し、累積文字画像数は約48,000点、木簡点数で約3,000点に達した。累計文字種も約1,500種で、木簡に登場する文字をほぼ網羅した。また、XMLの導入で意味による検索や他の情報とのリンクが可能になった。

共通入力ツールの開発 木簡データベースと木簡字典の共通入力ツールを開発し「木簡字典」のメタデータ付与を大幅に改善した。

東京大学史料編纂所「くずし字字典データベース」との連携検索の実現 2009年5月29日、東京大学史料編纂所とデータベース連携に関する覚書を交換。機関相互の画期的な連携が実現し、「木簡字典」と「くずし字字典データベース」との共通検索システムの開発に着手した。同年10月14日、完成した連携検索システムを公開。1000年にわたる字形の変化をカバーする検索が可能になった。

出土遺構年代観データベースと木簡人名データベース 木簡字典と MokkaShop を中核とした総合的な木簡研究拠点データベース構築のため、出土遺構年代観データベースを作成した。また、東京大学史料編纂所教授田島公氏を代表者とする学術創成研究の分担者として「木簡人名データベース」を開発し、出土遺構年代観データベースの出土地点情報を盛り込み、木簡字典とリンクさせた。

出土文字資料研究拠点の構築に向けて 本研究の究極目標は、木簡字典と MokkaShop を中核とした出土文字資料研究拠点データベースを構築し、木簡の整理・解読・保管のノウハウを形に残して社会に還元するとともに、それを研究工具として活用して新しい知見を再びノウハウに追加していく、いわば「知のスパイラル」を構築し軌道に乗せることにある。3年間の研究で、そのための扉を開くことができた。今後、木簡字典・MokkaShop を介した利用者との接点として、奈良文化財研究所のホームページで公開中の「木簡ひろば」の機能充実を図りたい。



5. 今後の計画

2011年度に、①木簡字典フラッシュ版の公開の実現、②墨書土器字典データベースの開発、③遠隔地システムの本格的な運用開始、2012年度に①木簡字典に外国語検索ページを用意し、海外との連携に備える、②知識データベース群の充実、③研究拠点データベースの本格的な運用、④印刷版『日本古代木簡字典』の改訂などを実現し、「知のスパイラル」を確立したい。

6. これまでの発表論文等(受賞等も含む)
渡辺晃宏、今後の平城京研究と木簡研究、上代文学 105, pp.22-40、2010

井上聡・馬場基、文字字形総合データベース作成の試み—電子くずし字字典データベースと木簡画像データベース・木簡字典の連携について—、人間文化研究情報資源共有化研究会報告集 1, pp.99-112、2010

Akihito KITADAI, Sherini SOMAYEH, Masaki NAKAGAWA, Hajime BABA and Akihiro WATANABE, Non-linear Normalization of Damaged Characters for Search Refinement (口頭発表あり)、Proc. 2nd China-Japan-Korea Joint Workshop on Pattern Recognition, 1, pp.185-188、2010

Jun TAKAKURA, Akihito KITADAI, Masaki NAKAGAWA, Hajime BABA and Akihiro WATANABE, Techniques to Enhance Images for Mokkan Interpretation, Proc. 12th International Conference on Frontiers in Handwriting Recognition, Vol. 1, No. 1, pp.358-362、2010、**ITESOFT best poster paper award** を受賞

Sherini Somayeh・末代誠仁・中川正樹・馬場基・渡辺晃宏、古代代木簡解読支援システムにおける字体検索の高性能化、人文科学とコンピュータシンポジウム論文集、2010-15, pp.27-32、2010

渡辺晃宏、日本古代の習書木簡と下級官人の漢字教育、高田時雄編『漢字文化三千年』(臨川書店刊)、pp.91-112、2009年

末代誠仁・齋藤恵・戸根康隆・石川正敏・中川正樹・馬場基・渡辺晃宏、古代木簡解読支援のための文字パターン検索、情報処理学会論文誌、50(4)、pp.1444-1455、2009

末代誠仁・中川正樹・馬場基・渡辺晃宏、デジタルアーカイブ検索機能を有する木簡解読支援システム、日本情報考古学会第26回大会講演論文集、Vol.6-2009, pp.66-77、2009

ホームページ等：

<http://hiroba.nabunken.jp/>：木簡ひろば
<http://jiten.nabunken.jp/>：木簡字典
<http://r-jiten.nabunken.jp/>：連携検索
<http://www.tuat.ac.jp/~koukai/gakuho/2010/504/topics2.html>：東京農工大のHP