

木簡・くずし字解読システム— MOJIZO— 専門家でも解読できない2割に挑戦

～木簡・くずし字解読システム— MOJIZO —とは～

「木簡・くずし字解読システム—MOJIZO—」は、奈良文化財研究所と東京大学史料編纂所が共同で開発した、画像から検索する（画像引き）文字画像データベースです。文字（画像）のクラ（蔵）の意味を込めて、MOJIZOと名付けました。

【内容】

- ・画像による検索
- ・類似した文字画像を表示
- ・木簡と古文書・古記録の字形・字体を収録

収録データ内容

奈良文化財研究所：主に奈良時代の平城宮・京跡出土の木簡の文字画像を収録

東京大学史料編纂所：奈良時代から江戸時代に至る前近代の日本史史料（古文書・古記録・編纂物など、紙に書かれた史料）の文字画像を収録

収録データ数

奈良文化財研究所 5,166 件 (2016年3月現在)

東京大学史料編纂所 24,296 件 (2016年3月現在)

*解析プログラムに収録されていて、現在表示可能な文字画像数です。
今後順次拡充していきます。

◆公開日：2016年3月25日（金）

*報道解禁 ラジオ・テレビ・インターネット 2016年3月25日（金）午後5時
新聞等 2016年3月26日（土）朝刊

◆URL：ht tp://mo.jizo.nabunken.go.jp/

【開発の経緯】

・釈読の困難さ⇒効率悪い

文字資料を釈読する際、墨痕は明確であっても、どう読めばいいかわからない場合があります。これまでは『五體字類』などの字典類で調べながら何とか読みを確定してきましたが、時間がかかる上、それでも読めないものがしばしばあります。

・文字入力できないため、検索できない⇒文字画像から検索できるシステムの必要性

異体字などの場合、文字入力できない場合があります。そこで、その使用例を調べようとしても、検索できません。文字画像から検索できるシステムの必要性を感じ、開発に着手し始めました。

・奈文研と史料編纂所との連携 + 桜美林大学末代誠仁氏のシステム開発

奈文研と史料編纂所は2009年に連携を開始して以来、前近代の様々な文字資料を解読するための文字画像を集めたデータベースを提供してきました（別添資料参照）。両機関が所蔵する文字画像データは日本史を研究する文字資料としては不可欠なものです。両機関の所蔵データを基本資料に、画像処理技術、文字認識研究の成果を応用して、文字画像から検索できるシステムの開発に成功しました。

【システムの特徴】

・従来のテキスト入力によらない検索システム

従来の文字（テキスト）から検索するシステムと違って、入力できない文字であっても検索できます。また、これまで文字が記号かわからなかったものも、見当が付くようになります。

・奈良時代から江戸時代までのさまざまな文字資料をカバー

奈文研と東大史料編纂所が機関の枠組を超えての連携していることで、古代から近世初頭までの1000年以上にわたる、木簡、古文書・古記録などさまざまな文字資料の字形・字体から抽出された類似画像を一度に閲覧できます。木簡の字と古文書・古記録の文字は資料としての性格、時代性が異なり、書体、書風、字体も異なります。したがって、解析したい画像の資料の性格や時代により、解析結果一覧からユーザーの視点で候補を選択することが可能です。

【効果・展望】

・釈読の効率アップに

これまで難解な字の釈読に時間がかかっていたのが、短時間で処理できるようになります。

・木簡やくずし字の解読をより身近に

くずし字や木簡の字と言えは難解なイメージがありますが、本システムを使うことで、より多くの人に親しんでもらえることを期待しています。

・専門家でも解読できない難しい文字に挑戦

既存システムを超える、専門家でも難しい字の解読に挑戦しています。

・記号など非文字資料の検索への可能性

文字コードによらない検索システムは、文字だけでなく、非文字資料となる記号などの意味づけや位置づけの解明にも繋がります。

【課題】

・認識率の向上（収録データの拡充）

・機能の充実と利便性の向上

*まだ発展途上にありますので、今後、システム本体や収録データの拡充にむけた努力を続けていきます。

【関係するプロジェクトと科学研究費補助金】

この研究およびシステム開発は、奈良文化財研究所（担当：都城発掘調査部史料研究室）と東京大学史料編纂所（担当：附属画像史料解析センタープロジェクト電子くずし字字典データベース開発プロジェクト（代表 久留島典子、井上聡・稲田奈津子・造藤珠紀・山田太造・和田幸大・宮崎肇））の連携によるものです。

また、下記の日本学術振興会の科学研究費補助金による研究成果の一部です。

- ①「木簡など出土文字資料の資源化のための機能的情報集約と知の結集」
（基盤研究（S）、2013～2017年度（予定）、研究代表者 渡辺晃宏、課題番号 25220401）
- ②「歴史的な文字に関する経験知の共有資源化と多面的分析のための人文・情報学融合研究」
（基盤研究（A）、2014～2017年度（予定）、研究代表者 馬場基、課題番号 26244041）
- ③「歴史知識情報のオープンデータ化にむけたスキームと情報利活用手法の再構築」
（基盤研究（A）、2014～2018年度（予定）、研究代表者 久留島典子、課題番号 26240049）
- ④「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」
（基盤研究（C）、2015～2017年度（予定）、研究代表者 末代誠仁、課題番号 15K02841）

お問い合わせ先

奈良文化財研究所史料研究室（担当：渡辺・馬場 TEL. 0742-30-6837）

東京大学史料編纂所古文書古記録部古記録第2室（担当：井上 TEL. 03-5841-8414）

木簡・くずし字解読システム — MOJIZO —

専門家でも解読できない2割に挑戦します

ユーザ
;解析したい
文字画像



取込み

木簡・くずし字
解読システム
— MOJIZO —

解析...

結果一覧
似ている画像を表示

上位8件画面→各機関100件まで表示する画面



奈文研

史料編纂所



リンク
それぞれのDBへ

奈文研
木簡字典DB

史料編纂所
くずし字DB



解析例

解析結果一覧

東京文化財研究所 東京大学文学部

木簡・くずし字解読システム

古文字の文字認識(一文字)から部分の文字認識が可能です。

足

操作手順: 足 → 脚 → 脚付延長画

解析する

東京文化財研究所 木簡データベース検索結果

脚付延長画(上)の検索結果

足 1801-11-1549	足 1802-11-1549	足 1803-11-1549	足 1804-11-1549	足 1805-11-1549	足 1806-11-1549	足 1807-11-1549	足 1808-11-1549

東京大学文学部編纂所 木簡データベース検索結果

脚付延長画(下)の検索結果

足 1801-11-1549	足 1802-11-1549	足 1803-11-1549	足 1804-11-1549	足 1805-11-1549	足 1806-11-1549	足 1807-11-1549	足 1808-11-1549

さらに表示(奈文研側)

東京文化財研究所 木簡データベース検索結果

脚付延長画(上)の検索結果

東京文化財研究所 木簡データベース検索結果の検索結果の表示結果です。

足 1801-11-1549	足 1802-11-1549	足 1803-11-1549	足 1804-11-1549	足 1805-11-1549	足 1806-11-1549	足 1807-11-1549	足 1808-11-1549
足 1809-11-1549	足 1810-11-1549	足 1811-11-1549	足 1812-11-1549	足 1813-11-1549	足 1814-11-1549	足 1815-11-1549	足 1816-11-1549
足 1817-11-1549	足 1818-11-1549	足 1819-11-1549	足 1820-11-1549	足 1821-11-1549	足 1822-11-1549	足 1823-11-1549	足 1824-11-1549

さらに表示(東大編纂所側)

東京文化財研究所 東京大学文学部

木簡・くずし字解読システム

東京大学文学部編纂所 木簡データベース検索結果

脚付延長画(下)の検索結果

東京大学文学部編纂所 木簡データベース検索結果の検索結果の表示結果です。

足 1801-11-1549	足 1802-11-1549	足 1803-11-1549	足 1804-11-1549	足 1805-11-1549	足 1806-11-1549	足 1807-11-1549	足 1808-11-1549
足 1809-11-1549	足 1810-11-1549	足 1811-11-1549	足 1812-11-1549	足 1813-11-1549	足 1814-11-1549	足 1815-11-1549	足 1816-11-1549
足 1817-11-1549	足 1818-11-1549	足 1819-11-1549	足 1820-11-1549	足 1821-11-1549	足 1822-11-1549	足 1823-11-1549	足 1824-11-1549

(参考) 奈良文化財研究所と東京大学史料編纂所との共同研究について

奈良文化財研究所

奈良文化財研究所は、1952年、文化財保護委員会に付属する文化財調査研究機関として設立されました。以来、古都奈良の文化財の総合的・学際的な調査・研究や、平城宮跡や藤原宮跡の本格的な発掘調査と研究を推進しています。こうした調査・研究の成果は、発掘調査報告書や研究書として公開しています。

また、全国の発掘調査に携わる自治体専門職員や海外の研究者を対象とした研修・共同研究を行うなど、埋蔵文化財を中心とした文化財調査・研究のナショナルセンターとしての機能も果たしています。ナショナルセンターとしての機能の一環として、文化財情報の積極的な公開に取り組み、上記報告書類の他、データベース群の公開や、データの交換標準に関する調査研究等も推進しています。

なお、今回連携の中心となるのは出土文字資料分野です。国内で出土文字資料の調査・研究に従事する専門チームを抱えているのは奈良文化財研究所だけです。日本の出土文字資料研究の中心拠点ということが出来ます。

【略歴】

- 1952年(昭和27年) 文化財保護委員会の付属機関として設置
- 1968年(昭和43年) 文化庁発足。その付属機関となる
- 2001年(昭和13年) 独立行政法人文化財研究所 奈良文化財研究所となる
- 2007年(昭和19年) 独立行政法人国立文化財機構 奈良文化財研究所となる

【奈良文化財研究所公開データベースの概要】

奈良文化財研究所ではナショナルセンター機能の一環として、さまざまな文化財情報のデータベースを公開しています。奈良文化財研究所が管理している文化財に関するデータベースの他、古都奈良の文化財の調査成果のデータベースや、埋蔵文化財全般に関するデータベース、さらには埋蔵文化財活用に関わるデータベースなども構築しています。公開データベースは23を数え、年間およそ66万件に及ぶアクセスがあります(2014年度実績)。

東京大学史料編纂所

東京大学史料編纂所は奈良時代から明治維新に至る前近代の日本史料(古文書・古記録・編纂物など、紙に書かれた史料)を組織的に集積・公開する国内唯一の研究所です。日本に関する史料であれば、国内外を問わず調査の対象とし、主として写真撮影による収集を進めています。

こうした活動を通じて築いた膨大な日本史料コレクション(複製史料約3万件、マイクロフィルム約490万コマ)を用い、編纂事業を行っています。1901年(明治34年)の『大日本史料』(『大日本古文书』)刊行以来、出版事業を継続し、現在その総発行数は1000点を越えるに至りました。また収集史料は、ご所蔵者の理解を得て図書室を通じ広く研究者・市民に対して閲覧公開されています。あわせて近年は電算機やネットワークの発展を受けて、史料情報や編纂情報に関する各種データベースの公開にも取り組んでいます。

【略歴】

- 1869年(明治2年) 和学講談所の修史事業を受けつぎ、史料編輯国史校正局が開設される
- 1888年(明治21年) 帝国大学(現東京大学)に移管され、臨時編年史料編纂所掛となる
- 1929年(昭和4年) 史料編纂所と改称

【東京大学史料編纂所データベースの概要】

東京大学史料編纂所は1984年以来、収集した史料に関する目録情報や、史料集編纂に関する各種データを広く公開する目的で、様々なデータベースを構築しています。対象とする史料は、奈良時代から明治維新にわたり、対象とする史料も古文書・古記録・編纂物・絵図・肖像画など多岐です。現在、公開データベースは32を数え、年間およそ350万件に及ぶアクセスがあります。

共同研究の概要

- ・日本を代表する拠点的研究機関の本格的な共同連携
 - ・2009年5月、両所長間で、奈良文化財研究所の木簡画像データベース『木簡字典』、史料編纂所『電子くずし字字典データベース』のデータベース連携に関する覚書や交換
 - ・2009年10月 「木簡字典・電子くずし字連携検索」公開
 - ・インターネット上により巨大な拠点を(バーチャルに)形成することに成功
- 両データベースが連結されることで、字画情報の検索範囲は、一挙にひろがりをもち、古代から近世初頭に至る字形の変遷を、一面のうちに鑑賞することが可能となった。以来、協力関係を強化しながら、現在に至っている。

◆奈良文化財研究所 木簡画像データベース『木簡字典』

飛鳥・藤原京、平城宮京出土木簡を中心とする木簡の文字画像データベース。2005年公開。検索したい文字を入力すると、カラー・モノクロ・赤外線画像・記帳ノートの画像を表示。文字検索の他、木簡の形状や大きさ、出土地からの絞り込み検索も可能。近年、関連する論文情報を表示する機能(2015年3月)や、意味内容から分類された語句を選択して検索できる意味検索機能(2016年2月)など、新しい機能も追加。また、外国語版(韓・中・英)も展開(2014年5月)している。

[収録データ内容およびアクセス数]

文字種	約 1,800 種
木簡点数(表裏別)	約 14,000 点
文字画像数	約 89,600 件
年間アクセス数	約 35,000 件 (以上、2015年度)

◆東京大学史料編纂所 『電子くずし字字典データベース』

所蔵する古文書・古記録に現れるくずし字の文字画像データベース。各種出典情報も付してデータベースを作成。単純に文字を入力してその字形を一覧するだけでなく、それを時代順に並べ替えたり、記主の属性(公家か武士かなど)に従って絞りこんだりすることも可能。また、くずした字形が似ている字を相互に参照する機能も付されている。

[収録データおよびアクセス数]

文字種/語彙種	約 6,000 件/約 10,000 件
字画画像総数	約 230,000 件
年間アクセス数	約 80,000 件 (以上 2015年度)