

記者会見 開催のお知らせ

木簡・くずし字解読システム—MOJIZO—の公開

1. 会見日時： 2016 年 3 月 25 日（金） 13：30 ～ 14：30

2. 会見場所：

奈良文化財研究所 多目的利用室（仮設庁舎 北棟 2 階）※別紙参照

3. 出席者：

奈良文化財研究所 史料研究室長	渡辺晃宏
主任研究員	馬場 基
研究員	高田祐一
東京大学史料編纂所古文書古記録部門 助教	井上 聡
桜美林大学総合科学系・専任講師	末代誠仁

4. 発表のポイント

◆出土木簡と古文書・古記録から収集した字形データをもとに OCR 機能を開発・公開し、任意の文字画像から誰でも似た字形を検索することが可能となった。

◆従来専門家に頼るほかなかった史料上の難読字形の解読について、誰もが一定の類推をすることが可能になった。また同機能は奈良文化財研究所・史料編纂所のホームページより誰でも無料で利用できる。

◆埋蔵文化財研究者や歴史研究者に役立つのみならず、社会教育・学校教育など幅広く活用されることが期待される。

5. 発表概要：

東京大学史料編纂所（以下、史料編纂所）と奈良文化財研究所（以下、奈文研）は、2009 年 5 月にデータベース連携に関する覚書を締結し、機関の枠組を越えた連携を行っている。2009 年 10 月に公開した「『文字画像データベース・木簡字典』『電子くずし字字典データベース』連携検索」は、1000 年以上にわたる日本の文字を簡易に検索できるデータベースとして、研究者のみならず、一般にも広く利用されている。

このたび、史料編纂所と奈文研では、これまでの文字から検索する（文字引き）連携検索システムとは別に、画像から検索する（画像引き）連携検索システム「木簡・くずし字解読システム—MOJIZO—」を開発した。文字による検索に加え、画像による検索も可能になり、文字が読めなくても似た字形を探しだすことが可能となる画期的なデータベース（9. 添付資料参照）となることが期待される。

6. 発表内容：

史料編纂所では古文書・古記録・編纂物に書かれた文字を読み解くため、奈文研では地中から発掘された木簡に記された文字を読むため、それぞれ字形・字体の画像を集めたデータベースを開発し公開を進めてきた。『電子くずし字字典データベース』『木簡画像データベース・

木簡字典』の両データベースは、対象とする資料は異なるものの、字を読むツールとしてのコンセプトや機能において共有する部分が大きく、2009年以來、機関の壁を越えて、連携検索を実現してきた。双方あわせて30万件をこえる字形画像と付随するメタデータを蓄積し、利用実績も年間10万件を越えるなど、学界を支える基本的インフラとして認識されている。

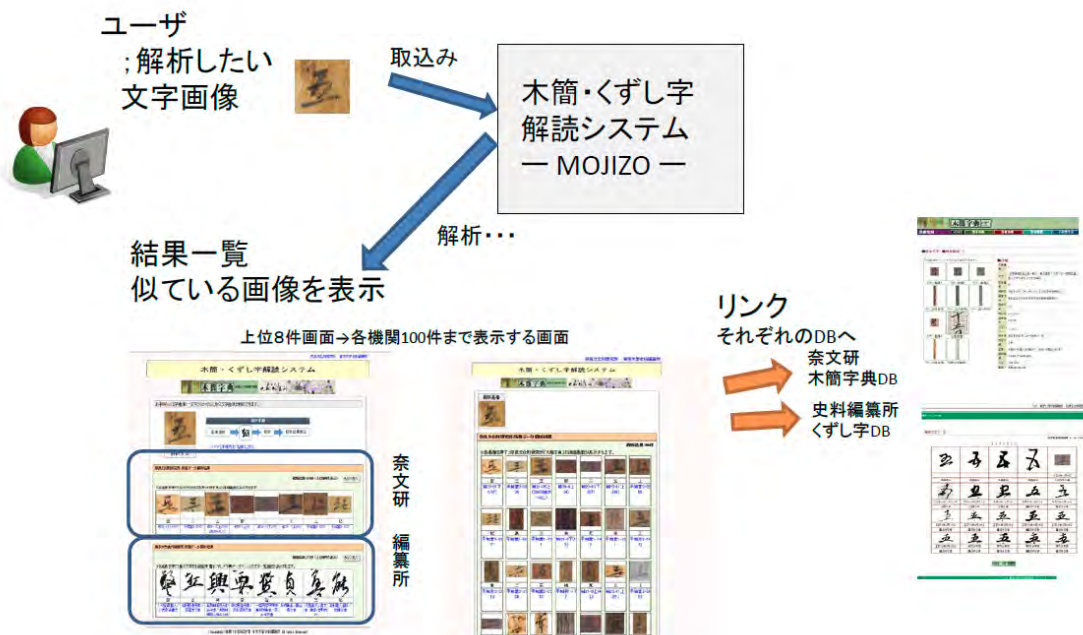
こうした蓄積をさらに多くの人々に利活用してもらうためには、字形やメタデータから検索するのみならず、実際に読めない文字画像を切り取ってシステムに投げかけ、蓄積データから類似のものを導き出す機能が必須となる。いわば歴史史料を読むためのOCRの必要性である。同様のOCR機能の開発は、さまざまな主体によって近年すすめられているが、典拠となるデータ群の大きさや確実性は、史料編纂所と奈文研が他機関よりも優れている。収集の対象も、古代から近代初頭に及び、木簡・石碑・文書・記録と典拠のバリエーション也多岐にわたっている。そこで両機関は、情報学研究者の援助を得て、典型的な字形3万余件を選んでさまざまに解析をすすめ、任意の字形画像に対して、近似するものを瞬時に、ネットワーク上で表示するシステムを開発した。同機能はMOJIZOとして、3月末より両機関のホームページにて無料公開する予定である。

MOJIZOの公開によって、文字史料は専門家以外の多くの人々にとって身近になることが期待される。身近にある石碑や古文書、軸物などの眺めるだけだった存在が、読み語る対象に変わっていき、文字史料は専門家に独占されるものではなく、生涯学習や学校教育のなかでも活用される貴重な素材になることが期待される。そうした関心が深まることで、遺跡や史料を保全する意識一般に広がってゆくならば、埋蔵文化財や歴史を研究する研究者にとっても大きな利益となることが期待される。

7. 添付資料：

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



(図)

画像から検索する（画像引き）連携検索システム「木簡・くずし字解読システム—MOJIZO—」を開発した。文字による検索に加え、画像による検索も可能になり、文字が読めなくても似た字形を探し出すことが可能となる画期的なデータベースです。