

木簡・くずし字読解システム「MOJIZO」再開のお知らせ

井上 聡

奈良文化財研究所との協力により、二〇一六年三月より公開となった標記システムは、史料編纂所歴史情報処理システムのリリースに伴う仕様変更のため、二〇二一年八月からしばらく連携停止を余儀なくされていた。多くの関係者のご協力を仰ぐことで、大変遅まきながら、ようやく本年三月より無事再開することができた（URL：https://amojizo.nabunken.go.jp）。以前と同様に活発なご活用を願う次第である。

「MOJIZO」は、ユーザーが入力した文字画像をもとに検索を行い、学習データとして格納する木簡の文字画像および古文書の文字画像から、それぞれ類似度の高いものを抽出し、順に並べて表示する機能を持つ。任意の文字画像があれば、その字が何であるか全く思いうかばなくても、候補を知ることができるところが大きな利点となっている。今回の連携復活を機に、要となる検索システムも、深層学習による特徴抽出をベースとするものへと転換され、解析精度の一層の向上が図られている。中核となるシステムの開発は、東京電機大学の大山航教授のご尽力によるもので、データの学習についてもご担当いただいた。

学習の対象としたデータは、本所の「電子くずし字字典データベース」(https://www.vaphiu-tokyo.ac.jp/ships/w34/search/character) が擁する代表字形約二万五千件と、奈良文化財研究所が公開する「木簡庫」(https://mokkankonabunken.go.jp/ja/) の代表字形一万五千余件で、「MOJIZO」の検索結果画面からは、それぞれの字形データベースへと遷移する。この基本設計は以前と変わらないが、「電子くずし字字典データベース」への遷移にあたっては、従前の問題を改良し、より機能的な接続を実現したところである。



図1 MOJIZO 検索画面

まだ改良の余地は多く残されているが、今後さらなる改善を重ねることで、ユーザーの期待に応えてゆきたい。ご利用のうえ、忌憚ないご批判をいただければ幸いである。

【付】奈良文化財研究所公開の「ナゾルくん」へのデータ提供

奈良文化財研究所が公開する市民参加型筆順情報取得ウェブアプリケーション「ナゾルくん」(https://nazorunabunken.go.jp/enter.html) に、本年三月から「電子くずし字字典データベース」のコンテンツを新規に追加公開した。同アプリは、市民や子供たちが、歴史的文字を「なぞり」書くことで、親しみを深めてもらうことを目的としており、情報生成の作業を共有した上で、筆順を当てるクイズで楽しんだりすることができ。ぜひアクセスいただき、お試しいただきたい。なお史料編纂所ホームページ上のH1ab(https://lab.hiu-tokyo.ac.jp) からリンクを示しているところである。



図2 MOJIZO 検索結果画面



図3 ナゾルくん フロント画面



図4 ナゾルくん 入力画面