

2022年10月26日

世界初のAIくずし字認識アプリ「みを (miwo)」が2022年度グッドデザイン賞を受賞
～くずし字教育や古文書による地域史料調査などにも貢献～

情報・システム研究機構 データサイエンス共同利用基盤施設 人文学オープンデータ共同利用センター（以下、ROIS-DS CODH）が2021年8月に公開した、世界初のAIくずし字認識アプリ「みを (miwo)」が、このたび2022年度グッドデザイン賞（主催：公益財団法人日本デザイン振興会）を受賞しました。

スマホ・タブレットアプリ「みを」は、約100万文字の「くずし字データセット」を学習した、最新のAIくずし字認識技術を用いたアプリです。昔のくずし字資料をいつでもどこでも読んで学べるよう、カメラで資料を撮影しボタンを押せば、AIが数秒でくずし字を現代日本語の文字に変換します。AIによるくずし字認識技術の進展を反映し、誰もがその技術の恩恵を受けられるように直感的に利用できるアプリとして公開したことが、グッドデザイン賞として高く評価されました。

「みを」は、Android版およびiOS版の両方がアプリストアにて無料でダウンロード可能です。アプリのダウンロード数はすでに約10万件、AIが認識した画像数も100万件に迫る勢いで増えています。

「みを」はさまざまな場所で使われています。くずし字を読めない方々が古文書などの内容を知るために使うだけでなく、くずし字に習熟した専門家も現地での下読み（迅速な内容確認）に活用しています。そして、教育機関における日本古典文学のためのくずし字学習、図書館・博物館・美術館などにおける資料内容の把握、市民による古文書の読み解きと地域の歴史の再発見などの目的に、アプリの利活用事例は拡大しています。AIが過去と現代を橋渡しすることで、文化遺産の利活用という社会課題の解決に新しい動きが生まれつつあります。

AIの支援により、多くの方々がくずし字を学び、古典を楽しむ世界の実現を目指して、ROIS-DS CODHは今後もアプリの改良を進めていきます。



受賞対象名：スマホ・タブレットアプリ [みを：AIくずし字認識アプリ]

事業主体名：ROIS-DS 人文学オープンデータ共同利用センター

プロデューサー：ROIS-DS CODH 北本朝展

ディレクター：ROIS-DS CODH 北本朝展 / Google Research カラーヌワット・タリン

デザイナー：Google Research カラーヌワット・タリン

受賞日：2022年10月7日

■ 関連ウェブサイト

グッドデザイン賞紹介ページ：<https://www.g-mark.org/award/describe/54368>



GOOD DESIGN AWARD
2022年度受賞

「みを」公式ホームページ：<http://codh.rois.ac.jp/miwo/>

■ アプリデザインのポイント

市民や専門家がくずし字資料をいつでもどこでも読んで学べるよう、重要な機能を選んでデザインしました。くずし字資料に関心を持つ方々の中には、必ずしもテクノロジーに慣れていないこともあり、操作マニュアルを読まなくてもアプリが使えるよう、シンプルなユーザーインターフェースとしました。

1. カメラ撮影画像の上にくずし字と現代日本語の文字を並べて表示することで、資料の内容を素早く確認できるようにしました。くずし字資料画像と現代日本語の文字表示とは、比較スライダーで切り替えられます。また文字を囲む四角形を表示することで、続けて書かれるくずし字（連綿体）の切れ目がどこにあるかも見えるようにしました。



2. アプリ利用者が AI くずし字認識結果を修正できるようにしました。AI の認識結果が適切ではないと思う場合、利用者が ROIS-DS CODH の「くずし字データセット」を参照して判断し、AI の認識結果を修正することで、より信頼性の高いテキストデータを作成できます。アプリの支援により、利用者がくずし字を読む能力も向上します。



■ アプリ開発の経緯

AI くずし字認識とは、深層学習（ディープラーニング）を用いて画像中のくずし字の位置と文字種を同時に認識し、くずし字を現代日本語の文字に変換する技術です。この技術を研究開発する過程では、ROIS-DS CODH のミッションであるオープンサイエンス(*1)の方針を重視してきました。以下にこれまでの主な研究成果をまとめます。

1. 2016年：人間文化研究機構 国文学研究資料館、情報・システム研究機構 国立情報学研究所と共同して、古典籍から収集したデータセット「日本古典籍くずし字データセット」を構築し、AI くずし字認識研究の基礎となる世界初の大規模データセットを、オープンデータとして公開しました(*2)。
2. 2018年：最初のAI くずし字認識モデルとなる KuroNet を開発し、IIF Curation Viewer を用いた「KuroNet くずし字認識サービス」を公開することで、AI を用いた多文字くずし字認識を世界で初めて実用レベルで実現しました。

- 2019年：世界最大の機械学習コンペプラットフォームである Kaggle において、人文系のテーマで世界初となる「Kaggle くずし字認識コンペティション」を開催し、世界中の AI 研究者・技術者の知恵を集めるオープンイノベーションを進めました(*3)。
- 2021年：いつでもどこでもくずし字認識が使えるよう、世界初の AI くずし字認識アプリ「みを」を無料公開しました。
- 2022年：「みを」がグッドデザイン賞を受賞しました。

■ アプリ開発の背景

くずし字をきちんと読める人は日本全体で 1 万人程度と言われますが、日本で大切に保存されてきた過去のくずし字資料は数億点にも達するため、大量の資料を読み解くには AI の助けが不可欠です。そこで、当時 ROIS-DS CODH 特任助教（国立情報学研究所 特任研究員）だったカラーヌワット・タリン氏（現 Google Research）が、自身の 10 年以上にわたる源氏物語研究で多数のくずし字資料に取り組んできた経験を踏まえ、AI くずし字認識モデルの開発を進めるとともに、市民や専門家がくずし字資料を読むのに便利な機能をアプリにデザインしました。

(*1) 学術論文と研究データがアカデミアおよび社会で広く共有され、幅の広い研究活動がオープンに行われることで、研究活動の加速化や、社会と緊密な連携の上に成り立つ問題解決が進み、学術活動が新しい次元（＝オープンサイエンス）に移行することが世界的に期待されています。

(*2) 江戸時代の文字の字形データセットを国文研との協働で構築／機械と人間の学習のためのオープンデータとして公開

<https://www.nii.ac.jp/news/release/2016/1117-1.html>

(*3) 【プレスリリース】「くずし字」の認識に世界の AI 研究者・技術者が挑戦

<https://ds.rois.ac.jp/post-3451/>

<国文学研究資料館 教授 山本和明氏のコメント>

わたしたちは日本の古い文献を読むためにいつも苦労してきました。専門家はくずし字辞典を手にし、調べ、なんとか解読していたのです。辞典をもたない一般の人はその手立てすら知りません。読めなくて当然ですね。「みを」の登場は、教育の現場のみならず人々の意識をも変えつつあります。なによりも直感的に操作できるのが良い。操作自体が楽しい。無料なのが良い。操作マニュアルを読むのは苦痛でしかありませんが、「みを」はそれも解放してくれました。様々な場所で、様々な機会に、わたしたちはくずし字で書かれたものに実は遭遇しています。「みを」を手にし、自ら進んで楽しく読んでみようとする方々が増えることを願ってやみません。

■今後の展開

「みを」のグッドデザイン賞受賞を機に、「みを」の新バージョンをリリースしました。まず AI くずし字認識の新しいモデル RURI を開発し、文字背景の色や模様が複雑な場合でも精度が向上しました。また、学習対象となる「くずし字データセット」を拡大することで、これまで認識できなかった一部の文字がアプリで新たに認識できるようになりました。

また AI くずし字認識モデル RURI の活用先として、市民参加型翻刻プラットフォーム「みんなで翻刻」との連携や、日本古典籍の大規模デジタル化プロジェクトにおける全文検索機能の実現などを進めます。こうした研究を通して、日本古典文学や日本文化を楽しむ人々を拓げるための活動に継続して取り組んでいきます。

【本件に関する問い合わせ】

大学共同利用機関法人 情報・システム研究機構
本部広報室
TEL:03-6402-6214 E-mail : koho@rois.ac.jp

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所
総務部企画課 広報チーム
TEL:03-4212-2164 E-mail : media@nii.ac.jp