News Release

TOPPAN

2022 年 9 月 13 日
凸版印刷株式会社

凸版印刷、AI-OCR で古文書を解読するスマホアプリを開発

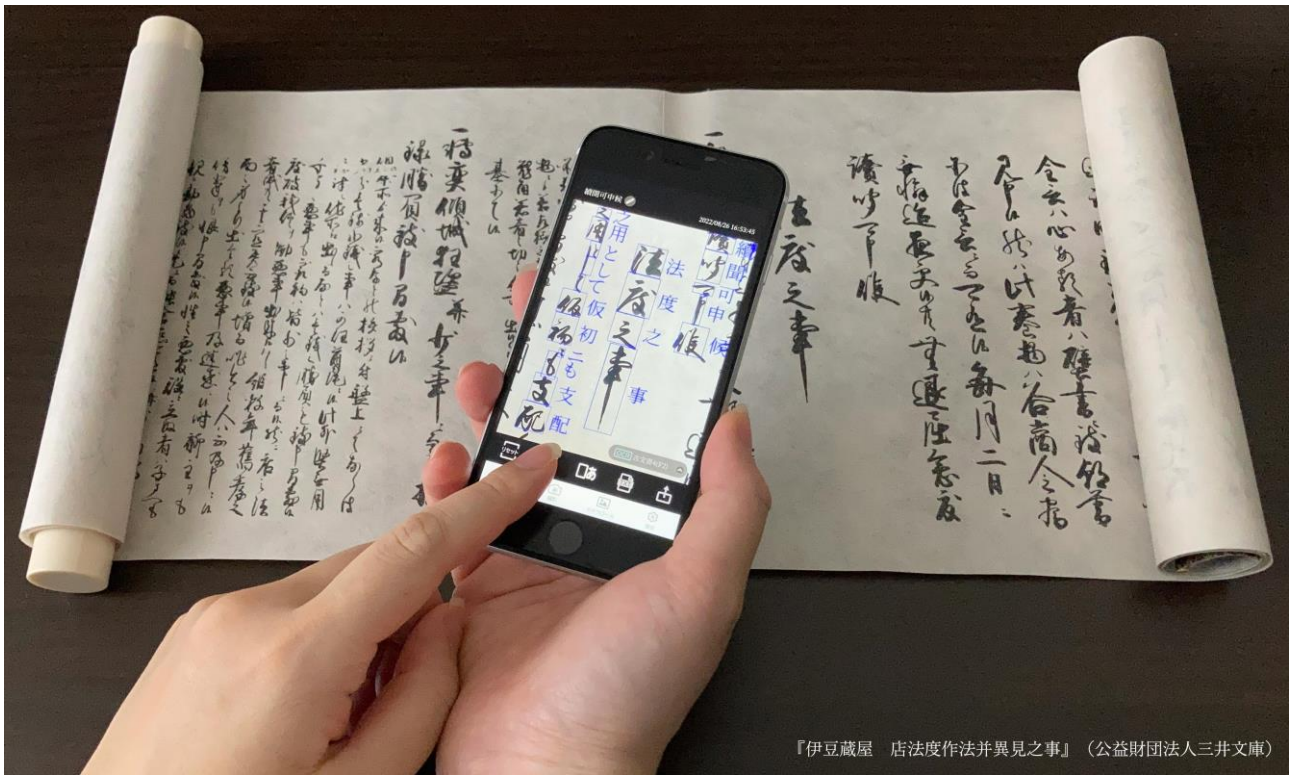
スマートフォンで撮影したくずし字資料を、その場で手軽に解読できるアプリケーションを開発。
資料館・大学等と連携し実証実験を開始、2023 年 3 月に正式リリース予定

凸版印刷株式会社（本社：東京都文京区、代表取締役社長：麿　秀晴、以下　凸版印刷）は、スマートフォンで撮影したくずし字資料を、その場で手軽に解読できるスマホアプリを開発しました。

2021 年にサービス提供を開始した古文書解読支援システム「ふみのは®ゼミ」（※1）がパソコン・タブレット上での利用、かつ法人向けに限られていたのに対し、本アプリケーションは一般利用者でもスマートフォンで撮影したくずし字資料を、その場で手軽に解読できるサービスです。

本アプリケーションは、木版を用いて印刷されたくずし字資料に対応した AI-OCR に加えて、新開発の手書きの古文書に対応した AI-OCR を搭載し、幅広い資料の解読を支援。資料館等での古文書の解読や調査業務の効率化をはじめ、一般利用者の「手元にある古文書の概要を知りたい」「くずし字を読めるようになりたい」といったニーズに対応いたします。

2022 年 9 月より公益財団法人三井文庫（東京都中野区、文庫長・武田晴人）、京都市歴史資料館（京都府京都市、館長：井上満郎）、和洋女子大学（千葉県市川市、学長：岸田宏司）などの協力のもと実証実験を開始。2023 年1月にベータ版公開、3 月に正式版の一般販売を予定しています。



古文書解読アプリ使用の様子

凸版印刷株式会社　広報本部
E-Mail:kouhou@toppan.co.jp

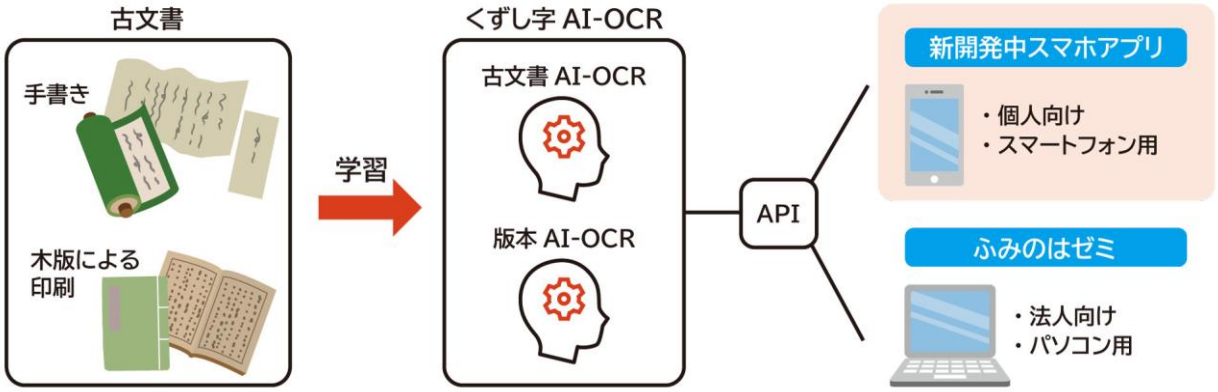
■ 開発の背景

日本国内に数十億点以上残存すると推測されている古文書には、循環型社会といわれる江戸時代の生活の様子や災害の記録といった、現代の社会課題にも直結する情報、また、地域特有の祭事や料理など、観光資源の創出や地域の活性化にもつながる貴重な情報が記されています。

しかし古文書のほとんどは「くずし字」で書かれているため現代人にとって判読が困難となってしまい、当時の記録・文献を活用する際の大きな障壁になっています。また、個人が所有している古文書は、内容がわからないために破棄されるケースも多く、解読されないまま災害による損傷や紛失、焼失などのリスクにさらされた状態で各地に眠っています。

凸版印刷は、これらの課題を解決する新たな手法として、2015 年より大学共同利用機関法人人間文化研究機構 国文学研究資料館との共同研究を開始し、以後、多数の研究機関等とくずし字 OCR 技術の開発・実証を重ねてきました。2017 年にリリースした原本画像と解読テキストを重ねて表示できる「ふみのは®ビューア」、2021 年にリリースしたオンラインくずし字解読支援システム「ふみのは®ゼミ」は、慶應義塾大学、早稲田大学坪内博士記念演劇博物館、大正大学をはじめ、多くの研究機関や大学などで採用されています。

手書きの古文書対応 AI-OCR は、公益財団法人三井文庫などの資料・データ提供協力の下、凸版印刷が独自に開発。また、「ふみのは®ゼミ」のリリース以降、「手元の古文書を手軽に読みたい」といった一般利用者向けのサービス提供について多数の要望をいただき、今回のアプリ開発に至りました。



「ふみのは」サービス全体像

■ 想定される利活用シーン

今回開発したアプリケーションは、専門家はもちろん、専門知識がない人でも利用が可能です。

研究機関や資料館等においてくずし字資料の事前調査・目録作りに本アプリケーションを使用することで作業の効率化を図ることはもちろん、個人の所有する古文書の解読を支援することで貴重な歴史資料の破棄や散逸の防止にも貢献します。AI-OCR を使うことで、これまでくずし字を学習したことのない人の「手元にある古文書の概要を知りたい」「くずし字を読めるようになりたい」などのニーズに対応します。

凸版印刷株式会社 広報本部
E-Mail:kouhou@toppan.co.jp

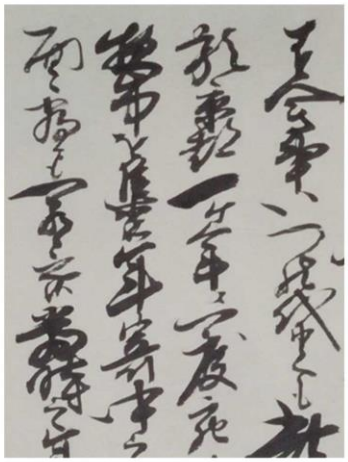
想定される利活用のシーン

■ 本アプリケーションの特長

・手書きと木版印刷物それぞれのくずし字資料に対応した AI-OCR エンジンを搭載。幅広い種類のくずし字解読に貢献

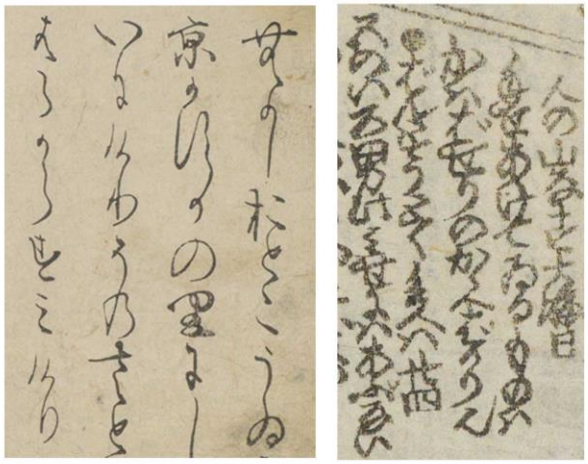
くずし字で書かれた資料は手書きのもの（書簡や証文、日記などの古文書）と木版を用いて印刷されたもの（版本や錦絵など）があり、それぞれ文字の形や使われている字種が異なります。本アプリケーションはそれぞれに対応した 2 種類の AI-OCR を搭載し、幅広い資料の解読を支援します。

新開発の古文書対応 AI-OCR は解読率 90%（※2）の精度を誇っています。



手書きの古文書の文字

『伊豆蔵屋　店法度作法幷異見之事』（公益財団法人三井文庫所蔵）



木版による印刷物の文字

左『伊勢物語』、右『大晦日曙草紙』（印刷博物館所蔵）

・AI におまかせの「フルオートモード」と、さらに高精度な「1 文字モード」が選択可能

「フルオートモード」は、画像の中にある文字領域を自動で検出し、つなげて書かれた文字の区切り位置も含めてAIがくずし字を解読します。さらに詳細に解読したい場合は、「1 文字モード」を使用することで、AIが提示する候補文字が表示されます。「フルオートモード」より高精度かつ利用者が文脈に合った文字を選択しながら解読することが可能です。

解読モードや AI-OCR の切り替えは、大きく見やすいボタンによって、パソコンやスマートフォンの操作に不慣れな方でも手軽な解読が可能です。

凸版印刷株式会社　広報本部
E-Mail:kouhou@toppan.co.jp

フルオートモード

画像の中の文字領域を自動で判別し、文字の区切り位置も合わせて解読。

１文字モード

利用者が読みたい範囲を指定して、AI-OCR を利用可能。AI が提示する複数の候補文字を表示する。

『伊豆蔵屋　店法度作法并異見之事』（公益財団法人三井文庫）

■ 「くずし字 OCR」技術について

OCR（Optical Character Recognition）とは光学文字認識のことで、文書画像に含まれる文字を読み取り、テキストデータに変換するソフトウェアの総称です。凸版印刷では 2013 年からさまざまな文献に対して、高い精度のテキストデータを提供する「高精度全文テキスト化サービス」を展開しています。このサービスで培ってきたテキストデータ化技術のシステム基盤を活用し、くずし字 OCR の研究・開発を進めてきました。2022 年 3 月にはくずし字認識コンペティションを開催するなど、日々さらなる技術向上に取り組んでいます。

凸版印刷、くずし字認識コンペティションを開催
https://www.toppan.co.jp/news/2022/03/newsrelease220322_1.html

■ 価格

販売価格：未定（実証実験の結果を踏まえ 2023 年 3 月までに決定）
※本サービスは法人向けの個別カスタマイズおよび API 提供も予定しています。

■ 今後の展開

本サービスは 2022 年 9 月より公益財団法人三井文庫、京都市歴史資料館、和洋女子大学などと実証実験を開始するとともに、iOS 版アプリは 2023 年 1 月にベータ版公開、3 月に正式版をリリースし App Store 販売を予定しています。

2025 年度までに、API 提供や関連事業を含め、一般利用者をはじめ、教育機関、博物館・資料館、地方自治体などへ向けてサービス提供を拡大し、約 3 億円の売上を目指します。

凸版印刷は本サービスをはじめ、全国各地に眠る貴重な歴史的資料の研究・活用の支援に継続して取り組んでいきます。

凸版印刷株式会社　広報本部
E-Mail:kouhou@toppan.co.jp

＜公益財団法人三井文庫　主任研究員　下向井 紀彦氏のコメント＞

現在くずし字解読システム「ふみのは®ゼミ」を使用した史料翻刻会を行っています。翻刻会は、①史料画像の全ページに一気に AI-OCR をかけて仮翻刻させる、②それをもとに AI の誤読・未読の文字を参加者で埋めていく、③穴埋めした翻刻文を使って内容を読み込んでいく、というやり方で進めています。過去に開催した史料翻刻会では、参加者が①部分を担っていたため、本務を抱える中での翻刻作業は負担でした。今回 AI-OCR で全文仮翻刻したため、参加者の作業を省力化でき文字修正と内容読解に注力することができました。

AI-OCR で、大量の史料の仮翻刻データをあっという間に作成できる意義は大きいと思います。例えば自治体史編さんや史料集刊行など、翻刻人材の不足している現場の負担軽減に寄与してくれるものと考えています。

他方、先日アプリの試作品に触れる機会を得ました。スマートフォン等のカメラで撮影した画像に AI-OCR をかけられるアプリで、保存しておいた画像に後から AI-OCR をかけることも可能です。出先での史料調査や史料リストの作成時に、史料の概要を把握する手助けになると思われます。また、調査現場で史料の撮影に専念して、後日 OCR をかけて内容確認する、といったこともできそうです。取り回しの良いアシストツールとして、現場での作業の省力化に繋げられるものと期待しています。

(※1)「ふみのは」サービスの詳しい説明についてはこちらをご覧ください。
公式ホームページ：https://www.toppan.co.jp/biz/fuminoha/

(※2) 古文書対応 AI-OCR は、近世の代表的な書体である御家流で書かれた資料を中心として字形を学習しています。精度 90%は御家流で書かれた古文書に対して AI-OCR を使って解読した際の結果です。

以　上

凸版印刷株式会社　広報本部
E-Mail:kouhou@toppan.co.jp