

2023年12月12日
TOPPAN 株式会社

TOPPAN、古文書解読スマホアプリ「古文書カメラ」のアップデートを実施

AI-OCR の改良と画像編集機能の追加により解読精度を大幅に向上。
2023年12月16日、17日開催のお城 EXPO 2023 などに出展

TOPPAN ホールディングスのグループ会社である TOPPAN 株式会社(本社:東京都文京区、代表取締役社長:齊藤 昌典、以下 TOPPAN)は、スマートフォンで撮影したくずし字資料を高精度の AI-OCR 技術によりその場で手軽に解読できるくずし字解読アプリ「古文書カメラ®」を、2023年6月より iOS 版、2023年10月より Android 版で提供しています。

「古文書カメラ®」は、江戸時代を中心とした古文書などのくずし字資料の解読を支援する誰でも無料で使用できるアプリケーションサービスです。

このたび、本サービスでは AI-OCR の再学習を実施し、解読精度の向上を実現するとともに、様々な条件下で撮影した画像に対応できるように、画像編集機能を搭載しました。また、解読機能の利用回数をこれまでの1日あたり10回から、30回に増加させています。

本アップデートは iOS 版、Android 版ともに 2023年12月12日に配信を開始します。



AI-OCR の再学習、AI モデルの刷新、画像編集機能による解読精度向上

また、本サービスは「お城 EXPO 2023」(12月16日および17日、場所:パシフィコ横浜ノース)にも出展が決定(※1)しているほか、大阪・関西万博に向けた「TEAM EXPO 2025」プログラムの一環である、「日本文化の継承と魅力発信セミナー&ワークショップ」(12月15日および16日、場所:グランフロント大阪北)(※2)における取り組み紹介が決定しています。イベントの様子やアプリの使い方は、「ふみのは®」の YouTube 公式チャンネルでの発信を予定しています。(※3)

TOPPAN は今後も、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していきます。

■ 開発の背景

古文書は、日本国内に数十億点以上残存すると言われていますが、現代の社会課題にも直結する災害や地域文化の記録など、防災や観光資源の創出・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う、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全国各地に眠っています。

TOPPAN 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 年に提供を開始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 年に提供を開始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」をはじめ、古文書解読とくずし字資料の利活用サービス「ふみのは®」として、様々なくずし字解読ソリューションを提供しています。（※4）

現在配信中の「古文書カメラ®」は、「ふみのは®」サービスの一環として TOPPAN が独自に開発したくずし字解読アプリです。

■ 主なアップデート内容

① AI-OCR の解読精度向上

学習データの追加および AI-OCR モデルの刷新により、とくに手書きの古文書に対して、大幅な解読精度の改善が認められました。

<解読精度向上について>

対象資料	6月版	12月版(今回)	精度改善
サンプル A (版本)	約 85%	約 90%	+5pt
サンプル B (古文書 読みやすいもの)	約 55%	約 84%	+29pt
サンプル C (古文書 読みにくいもの)	約 46%	約 69%	+23pt

※それぞれ 100 枚程度のサンプルに対して、レイアウト認識から文字認識までフルオート処理をした場合の精度を、正解テキストとの比較を行うことで計測

② 画像編集機能の搭載

アプリで撮影した画像や、スマートフォンから取り込んだ画像を編集することで、より AI-OCR の認識精度を高めることができます。



画像補正機能の「台形補正」と「ノイズ除去」のイメージ

<本アップデートで搭載した画像編集機能>

- ・台形補正:斜めから撮影した画像の補正
- ・ノイズ除去:シミやノイズなどを軽減
- ・トリミング:余分な画像の除去
- ・角度補正:資料画像の回転
- ・白黒反転:モノクロ状態で白黒を反転

③ AI-OCR の利用回数を 1 日 10 回から 1 日 30 回に増加

解読機能の利用回数をこれまでの 1 日あたり 10 回から、30 回に増やしました。

■ 「古文書カメラ®」概要

配信プラットフォーム	App Store／Google Play	
価格	・アプリダウンロード無料 ・AI-OCR による解読機能は1日ごとに 30 回まで無料 ※利用回数は毎日 0 時にリセットされます。 ※法人でのご利用については、下記サイトのフォームよりお問い合わせください	
公式 Web サイト	https://camera.fuminoha.jp/	
サービスに関するお問い合わせ先	上記公式 Web サイトのお問い合わせフォーム	
アプリダウンロード用 QR コード	Apple App Store (iOS) 	Google Play (Android) 
SNS サイト	・X 公式アカウント https://twitter.com/fuminoha_jp/ ・YouTube 公式アカウント ふみのはチャンネル https://www.youtube.com/@fuminoha	

※アプリのダウンロードはスマートフォンだけでなくタブレット端末でも可能です。

※撮影条件や古文書の保存状態などにより、「古文書カメラ®」のくずし字解読結果には誤りが含まれることがあります。内容についてはご自身でご確認ください。

■ 今後の展開について

本アプリケーションは、今後さまざまな時代および筆跡の文字を学習させていくことにより、解読可能な文書の種類を増やし、また、他のサービスと連携することにより利用シーンの拡大を行っていきます。

TOPPAN は本サービスを自治体や研究機関に向け広く提供することにより、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み、様々な史資料のアーカイブ化を通じ、社会・文化に貢献していきます。

■ 「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。昨今ではディープラーニングと呼ばれる AI 技術を用いて構築することが主流になっています。

「古文書カメラ®」のバックエンドで動作するくずし字 OCR は、ディープラーニング技術をベースに

TOPPAN デジタル(株)の研究開発部門が独自に開発したアルゴリズムにより実現されています。TOPPAN では今後もくずし字 OCR 技術の研究開発を進め、さらなる性能改善を図って参ります。

(※1)お城 EXPO 公式ホームページ:<https://www.shiroexpo.jp/>

(※2) 日本文化の継承と魅力発信ワークショップ『古文書からひも解く「江戸時代の日本酒文化」』についてはこちらをご覧ください。 <https://team.expo2025.or.jp/ja/event/679>

(※3) YouTube 公式アカウント ふみのはチャンネル:<https://www.youtube.com/@fuminoha>

(※4)「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.com/ja/joho/fuminoha/>

- * 「古文書カメラ」および「ふみのは」は、TOPPAN ホールディングス株式会社の登録商標です。
- * 「QR コード」は、(株)デンソーウェーブの登録商標です。
- * 「Apple」及び「App Store」は、Apple Inc.の商標です。
- * 「Google Play」及び「Android」は、Google LLC の商標です。
- * 「IOS」は、Cisco Systems, Inc.またはその関連会社の米国およびその他の国における登録商標または商標であり、ライセンスに基づき使用されています。
- * 本ニュースリリースに記載された会社名および商品・サービス名は各社の商標または登録商標です。
- * 本ニュースリリースに記載された内容は発表日現在のものです。その後予告なしに変更されることがあります。

以 上