

2022年11月11日  
凸版印刷株式会社

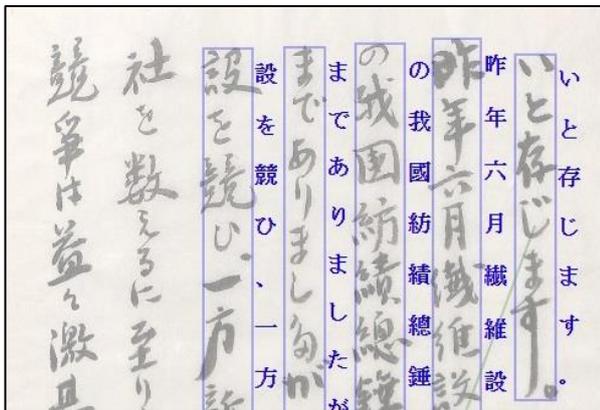
凸版印刷、明治期から昭和初期の手書き文字を解読する

AI-OCR を日本で初めて開発

くずし字を含む、多様な筆跡の手書き文字に対応する AI-OCR の開発に成功。  
大学や研究機関との実証実験を実施し、2023年4月よりサービス開始予定

凸版印刷株式会社(本社:東京都文京区、代表取締役社長:磨 秀晴、以下凸版印刷)は、明治期から昭和初期(以下近代)の手書き文字に対応した AI-OCR(※1)を開発しました。

近代の手書き資料には、くずし字で記されているものが多く、江戸時代の資料より解読が困難な場合もあると言われています。



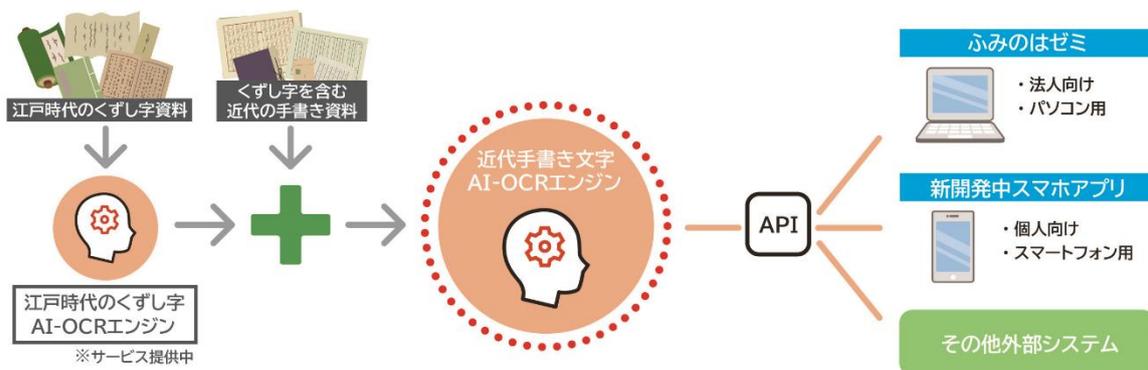
近代手書き文字 AI-OCR の解読例

(左) 鐘淵紡績株式會社『株主總會速記録』、(右) 鐘淵紡績株式會社『支配人回章』

※全て神戸大学経済経営研究所附属企業資料総合センター 所蔵

本 AI-OCR は、現在凸版印刷が古文書解読とくずし字資料の利活用サービス「ふみのは®」(※2)として提供している江戸時代のくずし字を対象とした AI-OCR をもとに、近代の多様な筆跡の手書き文字にまで対応の範囲を拡張させたものです。

今後、神戸大学経済経営研究所附属企業資料総合センター(所在地:兵庫県神戸市、センター長:西谷公孝)等との実証実験を2022年11月より実施、2023年4月より正式サービス開始を予定しています。



近代手書き文字 AI-OCR サービス全体像

## ■ 開発の背景

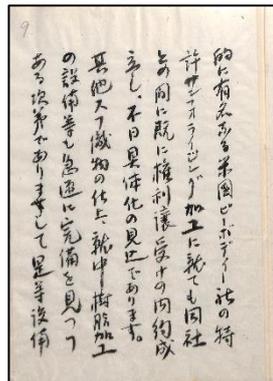
近代に作成されたくずし字を多く含む手書きの資料は、全国に大量に残存しており、その中には公文書、企業経営文書、業務日誌、書簡のような貴重な情報が記された資料が数多く含まれています。また、手書きの謄本類を取り扱う業務においては、近代の手書き資料の解読が日常的に行われていますが、これら資料の解読は、専門家にとっても難易度が高く手間のかかる作業であり、実用レベルの OCR 技術もこれまで開発されていませんでした。

このたび凸版印刷は、そのような課題を解決するため、神戸大学経済経営研究所附属企業資料総合センターおよび公益財団法人三井文庫(所在地:東京都中野区 文庫長:武田晴人)との共同研究のもと、学術研究や業務の効率化をはかるため、近代のくずし字を解読する AI-OCR 技術を独自に開発しました。

(神戸大学経済経営研究所附属企業資料総合センター所蔵)



鐘淵紡績株式会社  
『臨時株主総会決議録』

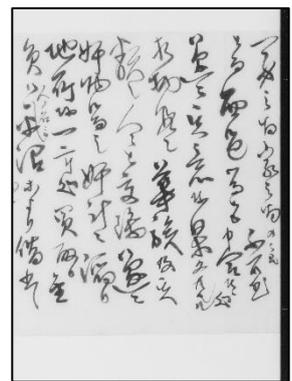


鐘淵紡績株式会社  
『株主総会速記録』

(公益財団法人三井文庫所蔵)



(旧)三井物産『日記』



『益田孝書簡(井上馨宛)』

近代手書き資料の例

## ■ 想定される利活用のシーン

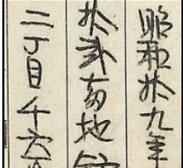
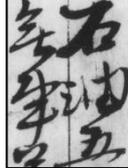
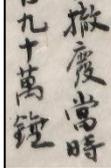
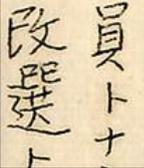
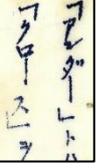
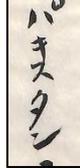
今回開発した AI-OCR は以下をはじめとした様々な利活用シーンを想定しています。

- ・ 地方自治体、教育研究機関、企業などが所蔵する難読資料の解読補助
  - ・ 難読資料の利活用を目的とした検索機能付きデータベースの構築補助
  - ・ 謄本類など日常的に難読資料の解読が必要な企業・機関の作業補助
- など

## ■ 本 AI-OCR の特徴

### ① 難読文字の解読

近代の手書き文字には、書き手によるくずし方のバラつきが大きい、筆記用具の多様化、カタカナ語が混在、旧字旧仮名遣い表記などの特性があり解読が非常に困難とされています。本 AI-OCR では様々なパターンの文字を AI に学習させることで、これらの難読文字の解読を実現しました。

難読化要因	文字の例		
書き手によるくずし方のバラつき			
筆記用具の多様化			
カタカナ語の混在			
旧字旧仮名遣い表記			

近代手書き資料 難読文字の例

## ②凸版印刷の複数のサービスと連携

サービス提供中のふみのは®ゼミシステムや、開発中の古文書解読スマホアプリ(※3)で提供予定です。

### ■ 「くずし字 AI-OCR」技術について

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、2015年以降、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はじめ、多数の研究機関とくずし字 AI-OCR 技術の開発・実証を進めてきました。現在では凸版印刷が提供するくずし字 AI-OCR は、古典籍や古文書を扱う業界において代表的な商用サービスとして認知されています。凸版印刷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](https://www.toppan.co.jp/news/2022/03/newsrelease220322_1.html)

### ■ 今後の展開

今後神戸大学経済経営研究所附属企業資料総合センターと2022年11月より実証実験を開始、2023年4月にふみのは®ゼミシステムの追加サービスとして正式リリースします。古文書解読スマホアプリ(開発中)の追加サービス、API形式での提供も順次リリース予定です。

2023年度のサービス開始初年度は売上2億円、2025年度には売上17億円を目指します。

### <神戸大学経済経営研究所附属企業資料総合センター長 西谷公孝氏のコメント>

近年、社会や環境を考慮したサステナビリティ経営の重要性が叫ばれています。しかし、近江商人の「三方よし」や渋沢栄一の「論語と算盤」から見て取れるように、サステナビリティ経営の考え方は、もともと古くから日本の経営理念に含まれていたものです。これは本来の経営のあるべき姿でもあるために、経営の歴史を紐解くことは、今後の持続可能な社会における経営のあり方を探る重要なヒントとなります。

一方で、それに着手しようとしても、これまで我々の前には近代手書き文字(くずし字)という高い障壁がありました。しかし、当センターとの共同研究の成果として今回開発された AI-OCR によって、専門家でなくともその解読ができるようになった結果、サステナビリティ経営、ひいては経営そのもののあり方の究明が飛躍的に進んでいく可能性があります。それほど今回開発された AI-OCR がもたらすポテンシャルは非常に大きいものです。この技術が様々な分野で利用され、「知」が探索されていくことを期待しています。

(※1) 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。本開発では人工知能(AI)を利用した OCR によって近代の手書き文字を解読します。

(※2)「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/>

(※3)凸版印刷、AI-OCR で古文書を解読するスマホアプリを開発

[https://www.toppan.co.jp/news/2022/09/newsrelease220913\\_1.html](https://www.toppan.co.jp/news/2022/09/newsrelease220913_1.html)

\* 「ふみのは®」は凸版印刷株式会社の登録商標です。

\* 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