News Release

TOPPAN

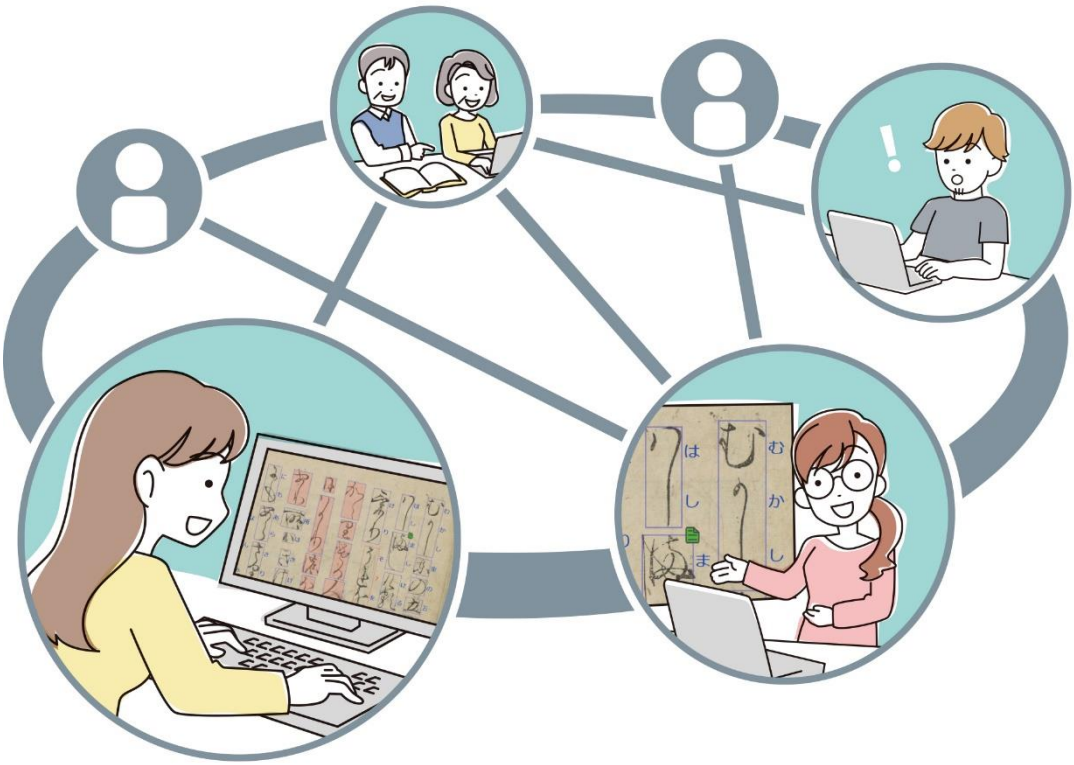
2021年 2 月 16 日
凸版印刷株式会社

凸版印刷、くずし字解読支援システム「ふみのはゼミ」を開発

高精度のくずし字 AI-OCR を搭載し、古文書・古典籍をオンライン上で簡単に解読できるシステム
共同作業をサポートし、コロナ禍における学習や研究、イベントなどに活用可能

凸版印刷株式会社（本社：東京都千代田区、代表取締役社長：鷹 秀晴、以下 凸版印刷）は、高精度のくずし字 AI-OCR を搭載し、古文書・古典籍をオンライン上で簡単に解読できるシステム「ふみのはゼミ」を開発しました。

授業やイベントでの活用を想定したグループワーク支援機能により、歴史的資料のデジタルアーカイブ化を推進するとともに、コロナ禍における学習・研究の拡大に貢献します。



「ふみのはゼミ」を活用した共同解読作業のイメージ

本サービスは、凸版印刷が 2015 年から研究・実証試験を行ってきたくずし字 OCR をさらに発展させ、凸版印刷総合研究所が開発した AI-OCR の導入による文字認識精度の向上および、グループワーク支援機能や、解読効率を向上させるためのさまざまなノウハウが結集したシステムです。パソコンやタブレットなどのブラウザ上で動作し、複数人での同時解読作業が可能になります。

AI-OCR 導入により、90%以上の高い精度で文字認識が可能（※1）となり、2019 年度には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（所在地：東京都立川市 館長：ロバート キャンベル、以下 国文学研究資料館）との実証試験を実施（※2）。また、慶應義塾大学と実験授業（※3）を行い、システムの改良に努めてきました。2020 年度には、慶應義塾大学をはじめとする 4 大学にて、オンライン授業内での演習用システムとしての活用を試験的に開始。教育機関や研究機関などでの利用を想定した機能のさらなる充実化を進めています。

凸版印刷株式会社 広報本部
〒101-0024 東京都千代田区神田和泉町 1 番地

■ 開発の背景

江戸時代以前に使用されていた「くずし字」は現代人には難読となってしまい、当時の記録・文献を解読する際の大きな障壁になっています。また、近年、大規模災害による資料アーカイブ手法の見直しや、資料の経年劣化や専門家の減少による文化継承の危機的状況などから、歴史的資料をデジタルデータとして保存することが求められています。

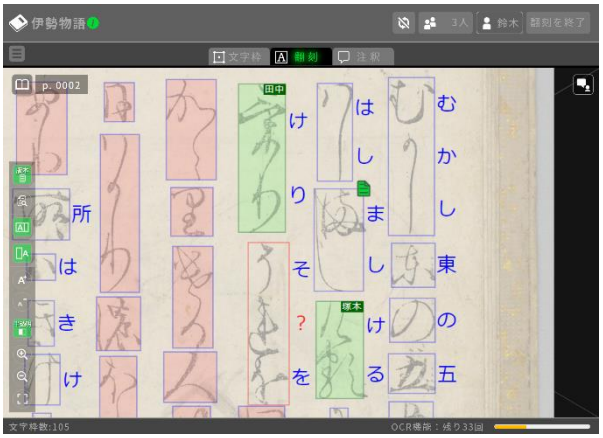
これらのニーズを解決する新たな手法として、凸版印刷は 2015 年より国文学研究資料館との共同研究により、くずし字 OCR 技術の開発・実証を重ねてきました。

今回開発した、くずし字解読支援システム「ふみのはゼミ」は、くずし字で書かれた歴史的資料が容易に読める環境を実現。また、オンライン上でのグループワークを可能にしたことで、コロナ禍における学習・研究等にも活用できます。

■ くずし字解読支援システム「ふみのはゼミ」の特長

・グループワーク用の支援ツール機能搭載

参加者が編集している箇所をリアルタイムで表示し、編集結果を即時反映する画面共有機能や、参加者同士で自由に交流できるチャット機能のほかに、解読した文字や単語に対して質問やコメントをつけることが可能です。講師への質問や、参加者同士の交流をスムーズに行うことができ、授業や各種イベント・ワークショップなどの活性化を促進します。また、授業やイベントでの利用を想定した、開始・終了の制御、採点機能なども搭載しています。

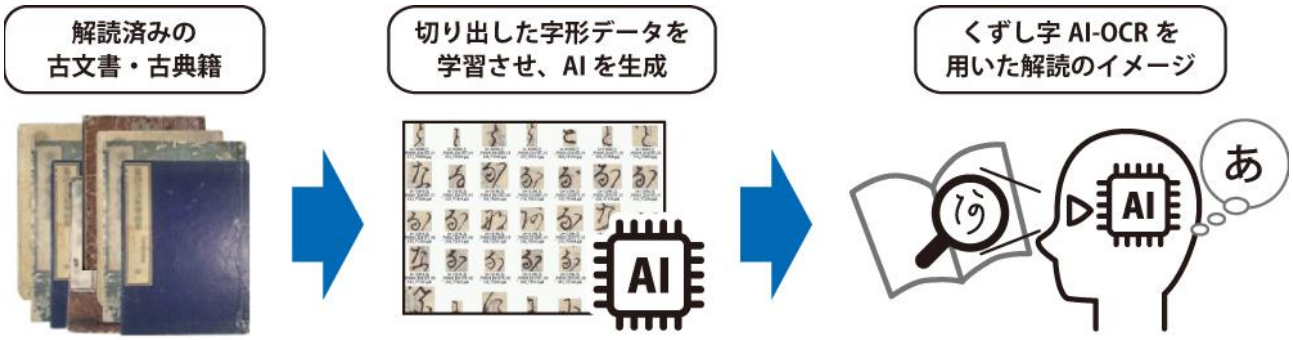


他の参加者が編集している箇所をリアルタイムで表示

（『伊勢物語』 印刷博物館所蔵）

・最新のくずし字 AI-OCR を搭載

解読済みの古文書・古典籍から字形を採集し、くずし字の形を AI に学習させることにより、AI-OCR を生成しました。「ふみのはゼミ」の画面上で、解読したい範囲を指定するだけで、AI が学習した大量の画像から、文字の区切り位置も含めて解読します。



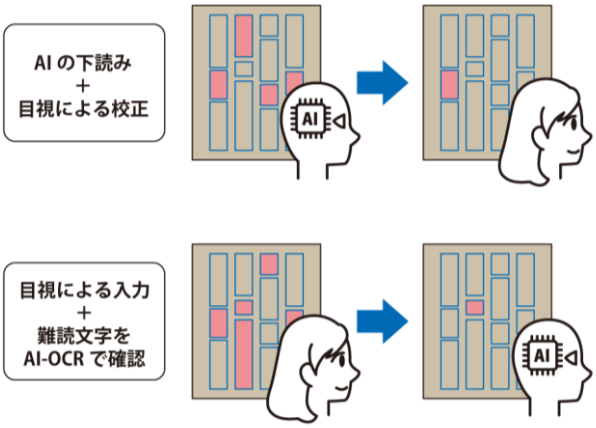
くずし字 AI-OCR 導入のイメージ

凸版印刷株式会社 広報本部
〒101-0024 東京都千代田区神田和泉町1番地

・目視と自動処理の併用による精度向上

目視による解読と、AI-OCR による文字認識の協調作業により、高い精度での解読を実現します。初心者から上級者まで練度に応じた方法で使用できます。

また、目視による入力・校正の結果を AI-OCR へ再学習させることで、AI-OCR の精度は向上していきます。



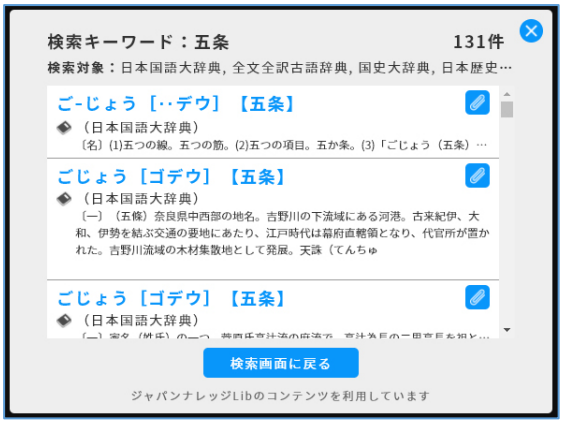
目視と自動処理のダブルチェックイメージ

・オンライン辞書・事典サイト「ジャパンナレッジ Lib」と連携

「ジャパンナレッジ Lib」が提供するオンライン辞書・辞典の検索 API と連携し、「ふみのはゼミ」からジャパンナレッジの辞書を検索することが可能です。文字を読むだけではなく、用例や背景を調べることにより、内容の理解を促進します。また、調べた用語を画面内にメモとして記録するクリップ機能を搭載しています。

・「ジャパンナレッジ Lib」公式ホームページ
https://japanknowledge.com/library/

※ジャパンナレッジ連携サービスの利用には別途「ジャパンナレッジ Lib」の契約が必要です。



「ジャパンナレッジ Lib」連携機能の検索結果イメージ

・既存のコンテンツとシステムの有効活用

所蔵資料の画像をもとに、独自の学習コンテンツの作成が可能です。また、既存の地域資料を利用したワークショップや翻刻会等の開催が容易になり、貴重史料の保全・解読活動を活性化します。

「ふみのはゼミ」は IIIF 形式に対応しているため、IIIF 形式で公開されている資料は、簡単な操作で解読を開始できます。

■ 「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。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。このサービスで確立したテキストデータ化技術のシステム基盤を応用し、くずし字 OCR の研究・開発を進めてきました。

2015 年にリリースした解読したテキストと原本画像を同時に表示できる「ふみのはビューア」は、早稲田大学演劇博物館をはじめ、多くの機関に導入されています。

凸版印刷株式会社 広報本部
〒101-0024 東京都千代田区神田和泉町１番地

■ 価格

【授業でのご利用】 10 万円～／半期（教育機関に限定し、週 1 回のご利用を想定）
【ワークショップ・イベント等でのご利用】 20 万円～／1 回
【翻刻会等でのご利用】 7 万円～／月額

※ご利用されるデータ容量や人数・利用形態によって価格は上下します。詳細は「ふみのは」の公式ホームページをご覧ください。

※お客さまが所蔵する資料から「ふみのはゼミ」で使用するデータを当社で作成する場合には、別途料金がかかります。

※講師やイベントスタッフの派遣、オペレーショントレーニング、機材貸与、イベントの企画等、には別途料金がかかります。

※料金は税別です。

■ 今後の展開

本サービスは教育機関、博物館・資料館、地方自治体などへ向け販売を開始し、2021 年9月までに一般利用に向けての開発を進めるとともに、2023 年までに関連事業を含め、約10億円の売上を目指します。

また、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・活用の支援に取り組んでいきます。

＜国文学研究資料館古典籍共同研究事業センター長　山本和明氏のコメント＞

30 万点にも及ぶ古典籍画像の公開を目指す国文研「歴史的典籍 NW 事業」は、くずし字で記された古典籍の全文テキスト化という壮大な夢の道半ばにいます。自然災害や感染症といった困難に直面する私たちには、参照すべき過去の記憶が、「記録」として今なお手付かずで残されているのです。

それを知りたいと願っても、残念なことに一部の人を除き判読するすべがありませんでした。今回開発された支援システムは、当館がオープンにした字形データなども活用され、オンライン上で解読可能な仕組みと伺っています。研究者のみならず学生や一般の人々にも、先人の知を開放することに繋がるもので、この事業に期待を込めてエールを送りたいと思います。

＜早稲田大学坪内博士記念演劇博物館　副館長　児玉竜一氏のコメント＞

コンピューターの力によって、くずし字を解読できる世が来るかもしれない。そんな話を初めて聞いたのは、前世紀の末のことでした。今や AI と手を携えながら学ぶことができる時代が本当に来ようとしています。

演劇の分野でも、特殊な字形で記される演劇資料の字形データベースから、AI による類推を交えて、原資料のくずし字読解授業などを試みています。まったくの初学者や、留学生たちにも、抜群の教育的効果があり、時間さえあえば海外の友人とも共同で、原資料を共有したくずし字セミナーが開けそうです。

研究の進展や資料の充実に伴って、くずし字読解の必要性は古典籍のみならず、近代の書簡や自筆原稿の世界にも拡がっています。既存のコンテンツをも縦横に活用した開放的なツールによって、古典籍やくずし字の世界の風通しがよくなればと願っています。

＜慶應義塾大学　経済学部教授　津田眞弓氏のコメント＞

近年注目される AI でくずし字を読む試みが、いつどういう形で使えるようになるのか知りたいと考え、2019 年に「ふみのはゼミ」を使った実験授業や、シンポジウムを行いました。現状、この種の試みで最も効果を発揮するのが教育での利用のようです。

AI と一緒に学習することは、特に初学者の教育に効果がありました。「ふみのはゼミ」はオンラインで動き、辞書データベースの連携や採点機能と、教育ツールとして進化しています。コロナ禍での試用でも通常の授業より判読結果や報告の精度が高まりました。

国際的な共同授業なども視野に、くずし字を学ぶのが難しい状況下の学習希望者に役立つツールになることを切に希望します。

凸版印刷株式会社 広報本部
〒101-0024 東京都千代田区神田和泉町1番地

(※1)2015 年、古典籍に対して 80%以上の精度での文字認識が可能であることを原理検証。2019 年の実証試験により、古典籍に関して自動処理の場合 90%以上、目視入力と補助的な OCR の利用の場合 95%程度の精度で解読可能であることを実証。古文書に対しては現在字形データベースの収集・分析と解読実証試験を並行して実施中。

(※2)くずし字OCR技術に関する実証試験結果:https://www.nijl.ac.jp/pages/cijproject/images/kuzushi-ji_ocr.pdf

(※3)慶應義塾大学教養研究センター実験授業『機械(マシン)と学ぶ「くずし字」』:http://user.keio.ac.jp/~sakura/kuzushiji_1/

＜「ふみのは」の詳しい説明についてはこちら＞

公式ホームページ:https://www.toppan.co.jp/biz/fuminoha/

＊ 「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

凸版印刷株式会社 広報本部
〒101-0024 東京都千代田区神田和泉町1番地