

自動認識システム大賞「優秀賞」

テーマ

古文書解読サービス「ふみのは®」

技術分野：OCR／画像認識

申請会社：TOPPAN デジタル株式会社／TOPPAN 株式会社

対象ユーザ：公的研究機関、大学、公文書館、地方自治体、一般利用者

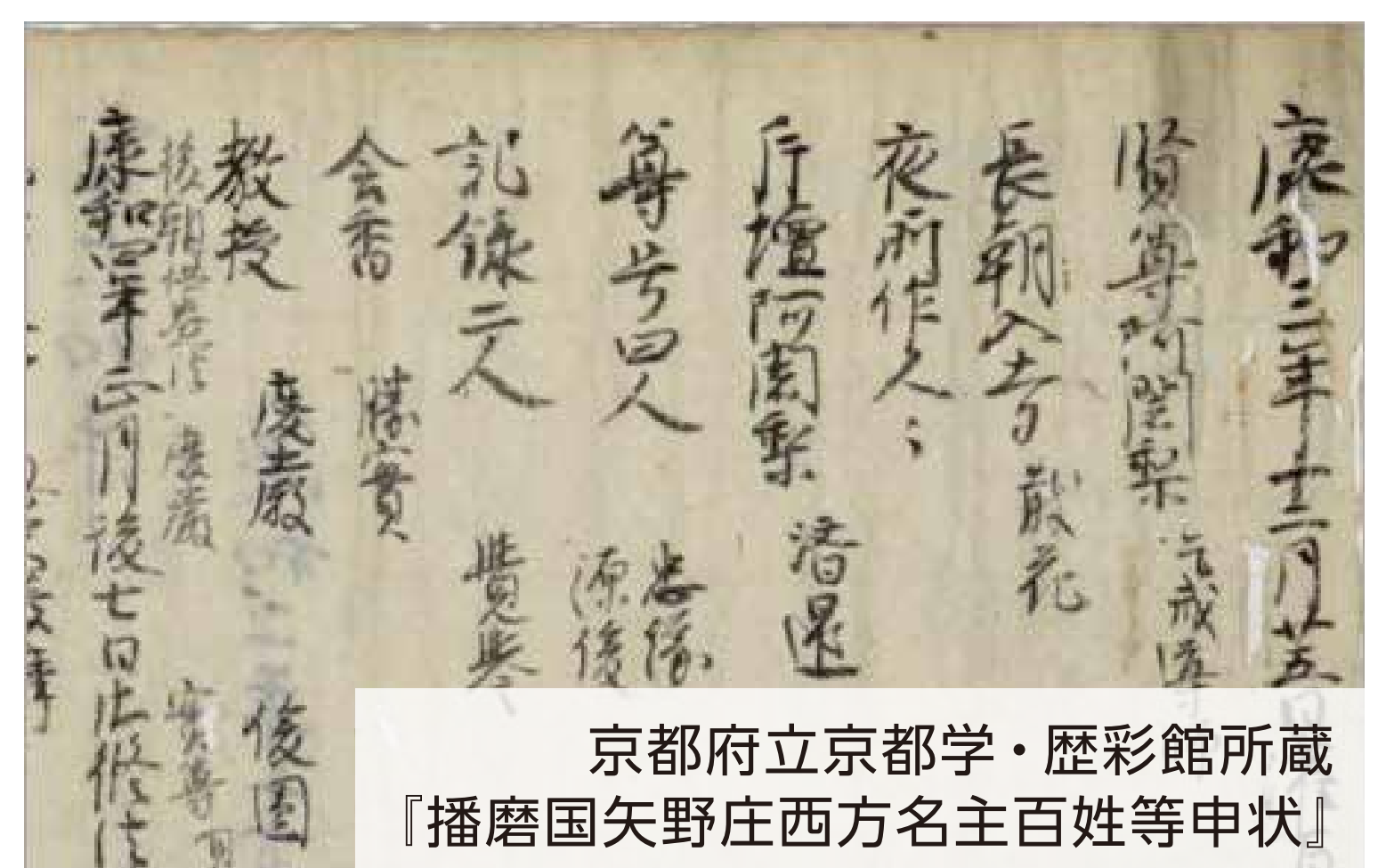
サービスの概要

国内に数十億点残存すると言われる古文書は、われわれに数百年前の社会や文化、災害などについて貴重な情報を与えてくれます。しかし古文書の多くは現代の日本人には解読が困難なくずし字で書かれており、解読できる人は 0.1% もいません。古文書認識サービス「ふみのは®」は、江戸期(まで)に書かれていたくずし字を文字認識し、翻刻を支援するサービスです。

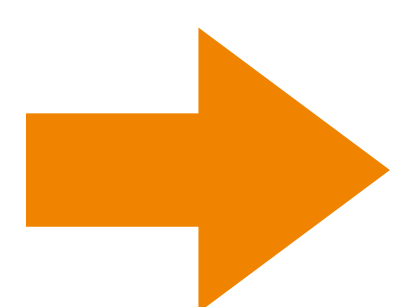
開発の背景

国内に数十億点残存すると言われる古文書(こもんじょ)

- ・数百年前の社会や文化、災害などについて貴重な情報を与えてくれる
- ・有効活用できれば大きな観光資源になり得るがその多くはあまり活用されずに各地に散在
- ・古文書の多くはくずし字で書かれており、解読できる人は 0.1% もいない
- ・古文書の内容が分からずに価値を判断できないために、破棄、劣化、紛失といった事象も起きている



京都府立京都学・歴彩館所蔵
『播磨国矢野庄西方名主百姓等申状』



現代人には解読困難になったくずし字を文字認識し、
翻刻(*)を支援するサービス「ふみのは®」を開発

※古文書・古典籍・石碑などに残された古い時代の文字を読み取りテキスト化すること

サービスの特長

多様なサービス形態

・ふみのはぜミ[®]【ASP 提供型】

くずし字 AI-OCR システムを PC ブラウザで動作するソフトウェアとして提供します。市民参加型のワークショップ、学校の授業、自治体の解読事業でご利用いただいています。



・古文書解読サービス【TOPPAN 請負型】

古文書の画像を TOPPAN がお預かりし、くずし字 AI-OCR システムを活用して解読した結果をご納品します。必要に応じて、識者による確認・検査を行います。

・古文書カメラ[®]【アプリ提供型】

一般向けのスマホアプリとして App Store/Google Play から配信しています。「その時、その場で」解読できることが大きな特徴で、効率的に資料の整理や調査を行えます。



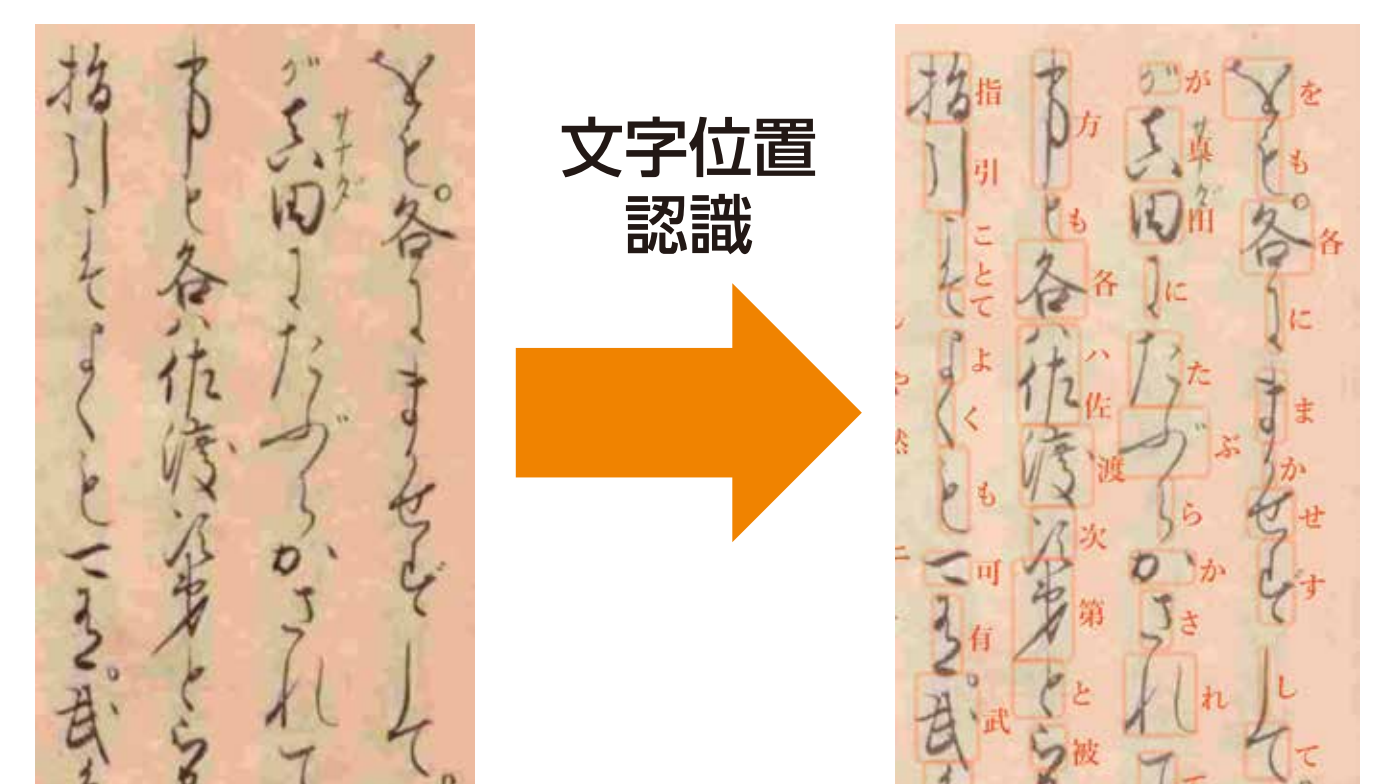
※「App Store」は Apple Inc. のサービスマークです。
※「Google Play」は Google LLC の商標です。

くずし字 AI-OCR 技術

・文字位置認識

くずし字のような難読文字で書かれた文書の解読においては、文書内に何が書かれているかだけでなく、文字ごとに「どこに」記載されているかを特定することが実用上重要になります。

文字と字が区切れなく繋がっていても、**文字位置**を認識。大学での授業や翻刻作業など、古文書に関わる人々を強力に支援します。

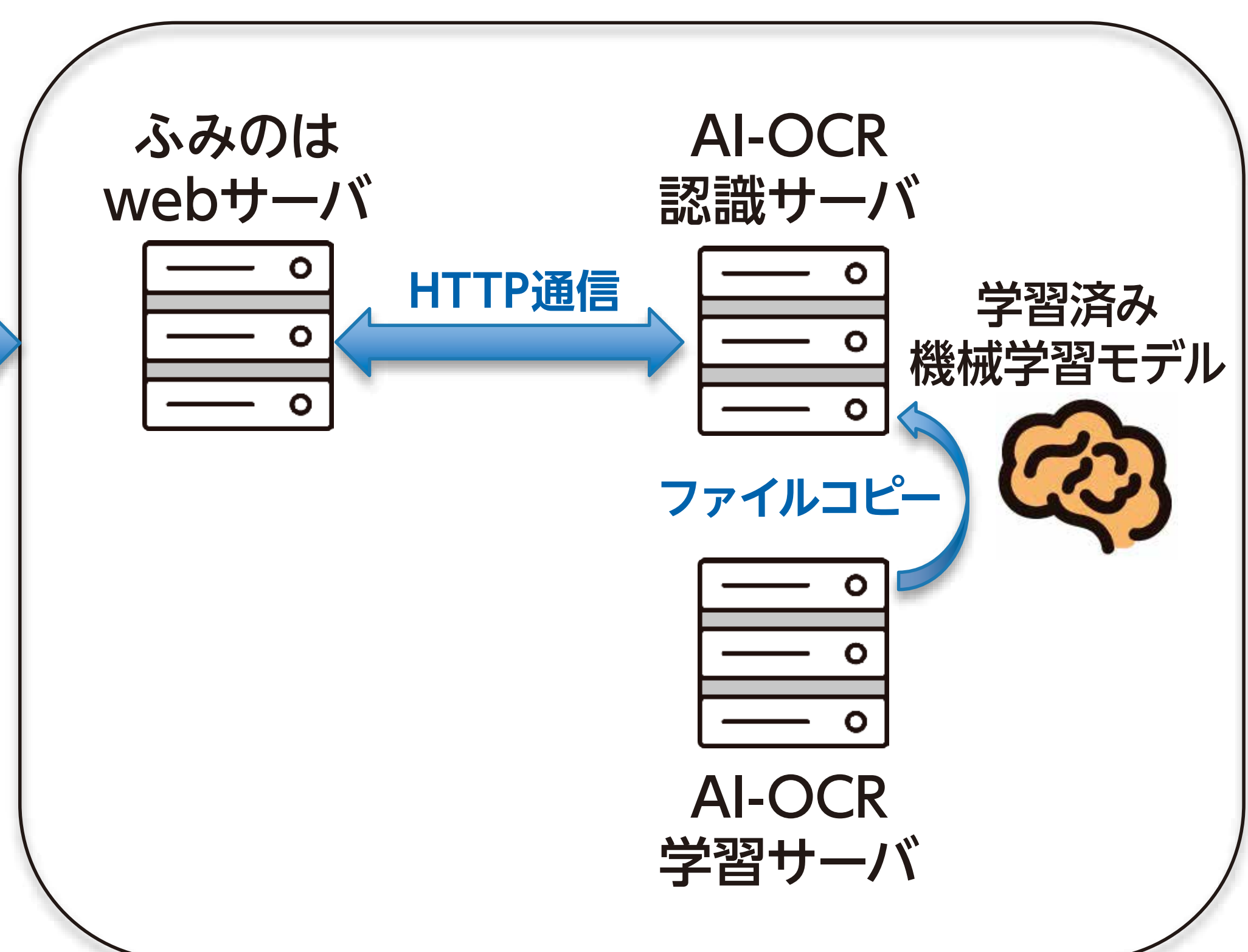


システム構成

利用者環境



ふみのはシステム



導入の効果

【経済的效果】

現時点では冒頭に掲げた数十億点の古文書の大規模な解読には至っていません

- しかしTOPPANグループでの技術開発の進展を受けて文字認識精度は着実に改善しており、学習データセットの拡充も進展するという、好循環が生まれています。

将来的に大規模な解読が実現された暁には、様々な経済的な効果が生まれるものと期待されます

【品質的效果】

約70～90%(F値)の認識精度

対象資料	認識精度
サンプル A (版本)	約 90%
サンプル B (古文書 読みやすいもの)	約 84%
サンプル C (古文書 読みにくいもの)	約 69%

導入事例

【解読業務への導入事例】

- ・ **郷土史家の寄託資料の整理・台帳作成業務** / 岐阜県垂井町様
郷土史研究家が収集した、約2万点の資料の調査・整理、目録作成業務の補助にふみのはをご活用いただきました。従来の目視のみの作業では3年以上かかると予想されていたのに対して、ふみのはの活用により、1年で作業を完了させることができました。
- ・ **国宝『尚家文書』** / 那覇市歴史博物館様
2023年度 沖縄振興一括交付金採択。1,292点ある**国宝『琉球国王尚家関係資料』**の内、『尚家文書』首里城に関する文書を3カ年で翻刻する計画です。尚家文書のレイアウトおよび字形、語彙をAI-OCRに学習させることで、解読精度の向上を実施しました。
- ・ **東北発 博物館・文化財等防災力向上プロジェクト** / 岩手県立博物館様
東日本大震災発生後、沿岸部が大規模な津波被害を受けた岩手県が中心となり、全国の専門機関が連携して被災した文化施設や収蔵資料等の再生に向けた取組を支援し、次の災害に備えて共有することを目的とした事業において、ふみのはゼミが採用されました。被災文化財のモデルケースとして陸前高田市の『吉田家文書』を岩手県内の広域ネットワーク(学芸員、教員、等)のサポートにより解読しました。

【ワークショップへの導入事例】

- ・ **連続講座「AIを活用した古文書ワークショップ」** / 京都市様
2021年度文化庁「地域文化財・普及啓発事業」にて実施。京都市の文化財保護などの団体に所属する一般市民を対象として、全5回開催しました。
- ・ **歴史謎解きゲームに挑戦！古文書探偵団** / 市立伊丹ミュージアム様
企画展「酒を醸す、酒をたしなむ」の展示史料を使った謎解きイベント。併設された酒蔵と展示室との回遊を目的として企画・謎の作成をいたしました。