

2022年3月22日  
凸版印刷株式会社**凸版印刷、くずし字認識コンペティションを開催**AI技術者の英知を結集し、画像から文字を認識する新しい技術を共創し、  
埋もれた歴史資料の有効活用を目指す

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、優秀な AI 技術者の知見を広く結集し、歴史資料の利活用に貢献するくずし字 OCR(光学文字認識)技術を高度化させるため、くずし字 OCR に関するコンペティション『くずし字認識チャレンジ』を開催します。本コンペティションは、株式会社 SIGNATE(シグネイト、本社:東京都千代田区、代表取締役社長:齊藤 秀、以下 SIGNATE)が提供しているデータサイエンスプラットフォーム「SIGNATE®」を活用し、2022年3月22日より募集を開始します。



くずし字で書かれた歴史資料『魯敏遜漂行紀略』(印刷博物館所蔵)

**■ 背景と狙い**

日本国内に数億点あるとも推測されている古文書や古記録などの歴史資料の多くはくずし字で書かれているため、現代では専門家以外には解読が困難になっています。一部資料は研究機関等で厳重に管理されていますが、内容が分からず体系だった管理もなされぬまま損傷や紛失、焼失などのリスクにさらされ各地に眠る資料も数多いといわれています。歴史研究の進展や文化継承の観点からは歴史資料の利活用を推進することが欠かせません。

このような状況を受け、凸版印刷は長年培ってきた OCR 技術を活用して 2021 年より古文書解読とくずし字資料の利活用サービス「ふみのは®」を提供しています。「ふみのは®」はこれまでに多くの研究機関への導入が進んでおり、更なる高度化、高精度化が望まれています。

このたび、SIGNATE が運用する国内最大規模の AI 開発技術者ネットワークを活用したコンペティションを開催することにより、凸版印刷がこれまで培ってきたくずし字 OCR 技術に新たな視点による最先端の AI 技術を取り入れ、くずし字 OCR 技術の飛躍的な向上を目指します。凸版印刷は社会的価値創造企業を目指し、文化資料の活用においても DX(デジタルトランスフォーメーション)を推進します。

## ■ コンペティション『くずし字認識チャレンジ』の実施概要

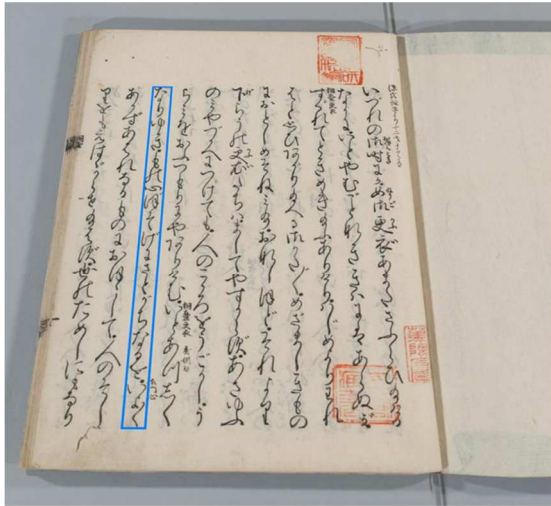
### ・テーマ

#### (A) 行領域認識部門

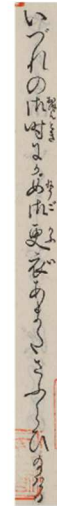
デジタルカメラなどで撮影した歴史資料を想定し、ページ画像から行領域を抽出する技術

#### (B) 行内文字認識部門

文字列を含む行画像を認識し、テキストシーケンスを出力する技術



(A) ページ画像から行領域(青枠)を認識



(B) 行画像についてテキスト認識

いづれの御時にか女御更衣  
あまたさふらひ給ける

『日本古典籍くずし字データセット』に含まれる『源氏物語』（絵入本、国文学研究資料館所蔵、doi:

10.20730/200003803、CC BY-SAにて配布、新日本古典籍総合データベースより）を加工した画像

左:行領域認識部門(版本を斜めから見るように再現する画像処理を施し、青枠を追加した画像)

右:行内文字認識部門(行ごとに切り取る処理を実施した画像)

### ・データセット

本コンペティションでは、データセットとして、『日本古典籍くずし字データセット』(国文学研究資料館ほか所蔵/情報・システム研究機構 データサイエンス共同利用基盤施設 人文学オープンデータ共同利用センター加工、doi:10.20676/00000340)を加工して利用します。元となるデータセットおよび加工されたデータセットはクリエイティブ・コモンズ 表示-継承 4.0 国際ライセンス(CC BY-SA)の下に提供されています。

国文学研究資料館 (<https://www.nijl.ac.jp/>)

人文学オープンデータ共同利用センター (<http://codh.rois.ac.jp/>)

CC BY-SA (<https://creativecommons.org/licenses/by-sa/4.0/>)

### ・インセンティブ

各部門選考の結果、1位から3位の入賞者には賞金を贈呈します。

### ・応募資格

SIGNATE 会員(登録無料)で、コンペティション参加規約に同意した方。

### ・応募方法・選考基準

SIGNATE ウェブサイト記載の内容をご確認ください。

(A) 行領域認識部門 URL: <https://signate.jp/competitions/580>

(B) 行内文字認識部門 URL: <https://signate.jp/competitions/581>

・募集期間:3月22日(火)~5月23日(月)

・選考結果公表:6月14日(火)

・主催:凸版印刷株式会社

## ■ 今後の展開

コンペティションの入賞技術を活用し、高精度な AI くずし字 OCR エンジンを開発します。またくずし字 OCR 技術開発で得られた技術・ノウハウを活用し、くずし字以外の特殊字形 OCR 技術の開発や BPO 事業事務局業務への AI-OCR 展開など、DX を実現する新規事業創出を目指します。

## ■ 「ふみのは®」について

凸版印刷が培ってきた OCR 技術を用いて、古文書や古典籍などくずし字で書かれた歴史的資料の利活用をサポートするサービスです。

<https://www.toppan.co.jp/biz/fuminoha/>

## ■ 「SIGNATE®」について

2022 年 3 月時点で約 63,000 人の AI/データ分析人材が登録する国内最大のデータサイエンスプラットフォームです。登録ユーザの 73%が社会人、27%が学生で、高度な知識やスキルを有した方々のコミュニティとなっています。

主な機能は、実社会のデータ分析課題に挑戦できる『SIGNATE Competition』、AI/データ分析の e-learning サービス『SIGNATE Quest』です。それぞれの登録ユーザに合った形で、企業や行政機関とのマッチングを通じた「成長」と「活躍」の機会を提供しています。

URL:<https://signate.co.jp/>

## ■ 「Erhoeht-X®(エルヘートクロス)」について

「Erhoeht-X®(エルヘートクロス)」とは、凸版印刷が全社をあげ、社会や企業のデジタル革新を支援するとともに、当社自体のデジタル変革を推進するコンセプトです。



「エルヘート」は、当社創業の原点である当時の最先端印刷技術「エルヘート凸版法」から名付け、語源であるドイツ語の「Erhöhen(エルホーヘン)」には「高める」という意味があります。

凸版印刷は、これまで培ってきた印刷テクノロジーの更なる進化とともに、先進のデジタルテクノロジーと高度なオペレーションノウハウを掛け合わせ、データ活用を機軸としたハイブリッドな DX 事業を展開し、社会の持続可能な未来に向けて貢献していきます。

※「SIGNATE」は、株式会社 SIGNATE の登録商標です。

\* 本ニュースリリースに記載された商品・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