

2023年9月27日

世界初・近代書籍対応の市販 AI-OCR ソフト 「FROG AI-OCR」新機能追加 ～旧字旧仮名遣いから新字新仮名遣いへの変換機能～

モルフォグループにおいて AI の事業化を担う、株式会社モルフォ AI ソリューションズ（所在地：東京都千代田区、代表取締役：神田武、以下 モルフォ AIS）は、世界初^{（注1）} 近代書籍対応の AI-OCR^{（注2）} ソフトウェア「FROG AI-OCR」に、旧字旧仮名遣いから新字新仮名遣いへの変換機能（例：旧字体「來」→新字体「来」）を新たに追加し、2023年9月27日（水）よりリリースします。



モルフォ AIS は、令和3年度から国立国会図書館の委託事業として図書館向け AI-OCR（NDLOCR）の開発に従事してきました。特徴としては、既存の OCR では読み取りが不可であった、近代書籍の旧字旧仮名や複雑なレイアウトへの対応、既存 OCR の2倍以上の読み取り精度^{（注3）}を達成いたしました。開発した AI-OCR プログラムは、NDLOCR として国立国会図書館が CC BY 4.0 ライセンスで公開しています。^{（注4）}

この度、図書館向け AI-OCR（NDLOCR）を活用した市販ソフト「FROG AI-OCR」（2022年7月1日発売）に新機能を追加しました。

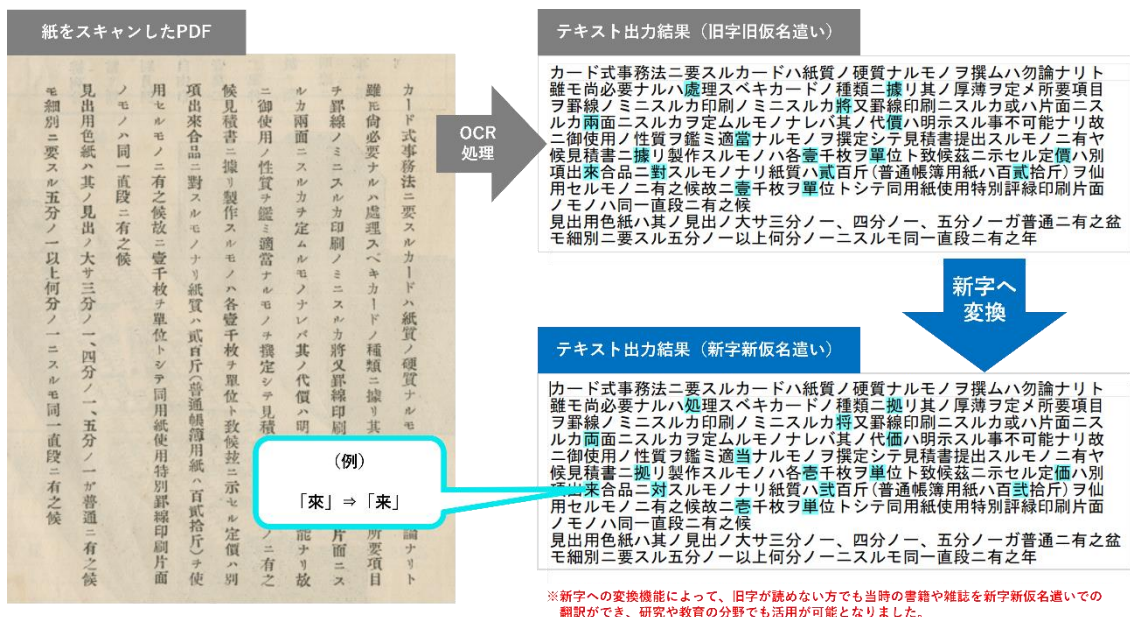
【新機能について】

■ 旧字旧仮名遣いから新字新仮名遣いへの変換機能

今回のアップデートにて、「旧字旧仮名遣い」から「新字新仮名遣い」への変換をすることができるようになりました。例えば、旧字体である「來」を新字体の「来」に変換することができます。

この変換機能を利用することで、旧字旧仮名遣いが読めない方でも、当時の書籍や雑誌を新字新仮名遣いで翻訳や検索することができ、研究や教育の分野でもより幅広い方に活用いただくことが可能となります。

例えば、透明テキスト付 PDF で新字新仮名文字を重ね合わせることで、過去の歴史的資料の検索性が圧倒的に高まると考えております。



【FROG AI-OCR 製品ページ】

<https://frog-ai-ocr.morphoai.com/>

【FROG AI-OCR 紹介ムービー】

<https://www.youtube.com/watch?v=MxAAHdQk4I>

【お申込み・問い合わせ窓口】

<https://www.morphoai.com/ai-ocr>

こちらより無償トライアル頂く事が可能です

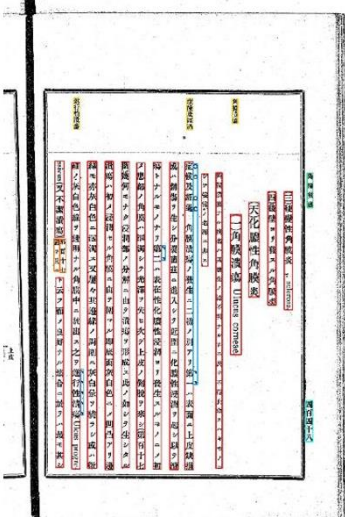
【新機能版リリース日】

2023年9月27日(水)

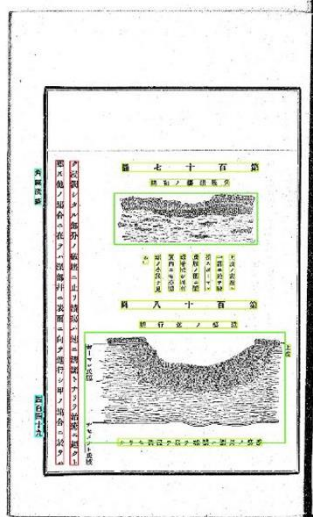
【NDLOCR の特徴】

① 複雑なレイアウトへの対応

本文・頭注・割注・ルビ・傍点・柱・ノンブル



本文・図表・キャプション・柱・ノンブル

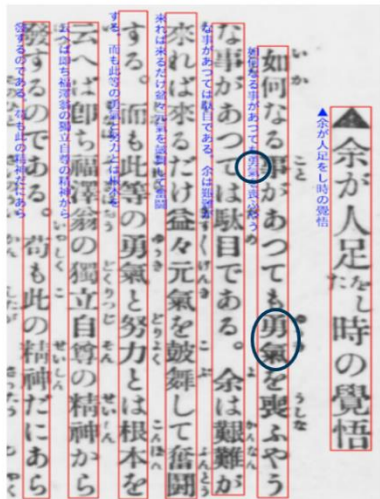


本文・多段組・割注・ルビ・柱・ノンブル

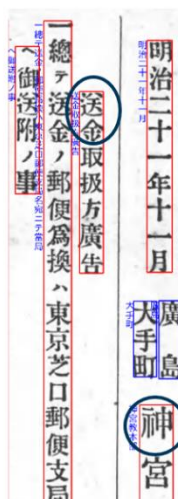


② 多様な文字種（旧字旧仮名）への対応

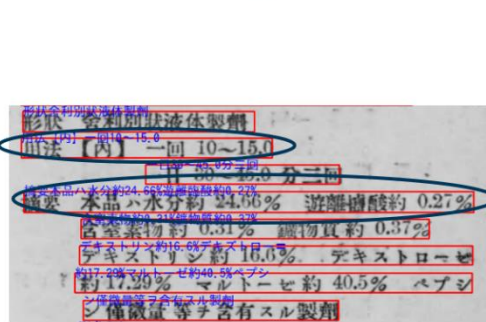
旧字体の認識



異体字→新字への変換



記号（〇、～、%）や英字数字

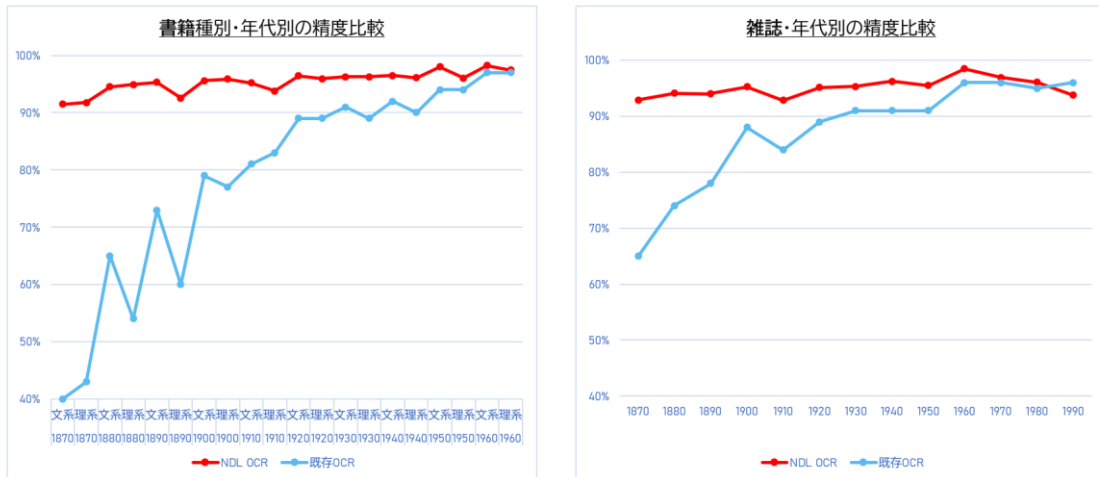


③ OCR 処理プログラムの精度向上

1860 年代以降の書籍・雑誌において市販 OCR より高い 90%以上の精度で判定可能となりました。

(注 5) 特に明治期～昭和初期の近代書籍・雑誌においては、市販 OCR より約 2 倍 (約 40%→90%以上) の読み取り精度となりました。

NDL OCRと既存OCRの精度比較(文字単位での一致度(F値))



(出典) 国立国会図書館 NDLラボ「令和3年度OCR処理プログラム研究開発」https://lab.ndl.go.jp/data_set/ocr/r3_morpho/

④ 透明テキスト付 PDF 出力機能

テキストファイルでの出力に加え、透明テキスト付 PDF での出力も可能です。こちらにより、出力した PDF 内でテキスト検索などを行うことが可能です。

紙をスキャンしたPDF ※テキストを選択できない

そうして驚愕の威力に手をかけながら、天候に老婆の前へ歩いた。老婆が驚いたのはいうまでもない。老婆は、一目下人を見ると、まるで「野」にでもはじかれたように、飛び上がった。「おのれ、どこへ行く」

下人は、老婆が死骸につまずきながら、あわてふためいて逃げようとする行手をふさいで、こののした。老婆は、それでも下人をつきのけて行くとする。下人はまた、それを行かすまいとして、押しもどす。二人は死骸の中で、しばらく、無言のまま、つかみ合った。しかし静寂は、はじめからわがわがしている。下人はとうとう、老婆の胸をつかんで、無理にそこへねじ倒した。ちよとど、胸の胸のような、骨と皮ばかりの胸である。「何をしていた。言え。言わぬと、これだぞよ」

下人は、老婆をつき放すと、いきなり、威力の精を払って、白い顔の色をその眼の前へつきた。けれども、老婆は黙っている。両手をわなわなふるわせて、鼻で息を切りながら、眼を、前後が顔の外へ出そうになるほど、見開いて、地のように軌跡で黙っている。これを見ると、下人ははじめて明白にこの老婆の生死が、全然、自分の意志に支配されているということ意識した。そうしてこの意識は、今までけわしく燃えていた憤怒の心を、いつの間にかまじってしまった。あとに残ったのは、ただ、ある仕事をして、それが円満に終わった時の、安らかな得意と満足とがあるばかりである。そこで、下人は、見下しながら、少し声を落してこう言った。

「君は極非道徳の『狂』の役人などではない。今、力この門の下を通りかかっ

透明テキスト付PDF ※テキストを選択できる!

そうして驚愕の威力に手をかけながら、天候に老婆の前へ歩いた。老婆が驚いたのはいうまでもない。老婆は、一目下人を見ると、まるで「野」にでもはじかれたように、飛び上がった。「おのれ、どこへ行く」

下人は、老婆が死骸につまずきながら、あわてふためいて逃げようとする行手をふさいで、こののした。老婆は、それでも下人をつきのけて行くとする。下人はまた、それを行かすまいとして、押しもどす。二人は死骸の中で、しばらく、無言のまま、つかみ合った。しかし静寂は、はじめからわがわがしている。下人はとうとう、老婆の胸をつかんで、無理にそこへねじ倒した。ちよとど、胸の胸のような、骨と皮ばかりの胸である。「何をしていた。言え。言わぬと、これだぞよ」

下人は、老婆をつき放すと、いきなり、威力の精を払って、白い顔の色をその眼の前へつきた。けれども、老婆は黙っている。両手をわなわなふるわせて、鼻で息を切りながら、眼を、前後が顔の外へ出そうになるほど、見開いて、地のように軌跡で黙っている。これを見ると、下人ははじめて明白にこの老婆の生死が、全然、自分の意志に支配されているということ意識した。そうしてこの意識は、今までけわしく燃えていた憤怒の心を、いつの間にかまじってしまった。あとに残ったのは、ただ、ある仕事をして、それが円満に終わった時の、安らかな得意と満足とがあるばかりである。そこで、下人は、見下しながら、少し声を落してこう言った。

「君は極非道徳の『狂』の役人などではない。今、力この門の下を通りかかっ

透明テキスト化

キーワード検索も可能!
(例) 「老婆」

※PDF画像上に透明なテキストデータを付与することで、「コピー&ペースト」や「検索」などが可能となります。

【注釈】

- (注 1) 弊社調べ、明治～昭和初期の書籍（文系・理系）及び雑誌の認識精度 90%以上を保有する市販の AI-OCR 製品が対象
- (注 2) OCR (Optical Character Recognition) 光学式文字認識
- (注 3) 令和 3 年度 OCR 処理プログラムの研究開発リリース参照
https://www.morphoinc.com/news/20220428-jpr-mais_ndl
- (注 4) NDLOCR のアプリケーション (GitHub 上で公開)
https://github.com/ndl-lab/ndlocr_cli
- (注 5) 国立国会図書館 NDL ラボ「令和 3 年度 OCR 処理プログラム研究開発」
https://lab.ndl.go.jp/data_set/ocr/r3_morpho/

【関連プレスリリース】

2022 年 4 月 28 日

モルフォ AI ソリューションズが、国立国会図書館の最新 AI 技術を活用した OCR 処理プログラムの開発を完了
https://www.morphoinc.com/news/20220428-jpr-mais_ndl

2022 年 5 月 9 日

モルフォ AI ソリューションズが、国立国会図書館から視覚障害者等用の OCR 開発を受託
https://www.morphoinc.com/news/20220509-jpr-mais_ndl

2022 年 6 月 14 日

世界初、近代書籍対応の市販 AI-OCR ソフト「FROG AI-OCR」新発売
https://www.morphoinc.com/news/20220614-jpr-mais_frog_aiocr

2023 年 3 月 27 日

世界初、近代書籍対応の市販 AI-OCR ソフト「FROG AI-OCR」新機能追加
～透明テキスト付 PDF 出力機能・多段組対応機能～
https://www.morphoinc.com/news/20230327-jpr-mais_frog_aiocr

【株式会社モルフォ AI ソリューションズについて】

モルフォ AI ソリューションズは、AI（人工知能）の事業化に取り組む企業です。行政、電力、交通、製造といった社会インフラの領域で、AI-OCRをはじめとする最先端の AI 技術の導入と実運用を推進しております。

所在地：東京都千代田区神田錦町 2-2-1 KANDA SQUARE 11 階 WeWork 内

代表者：代表取締役 神田 武

設立：2019 年 12 月

事業内容：AI コンサルティング、システムインテグレーション、SW・HW 販売など

ホームページ：<https://www.morphoai.com>

【株式会社モルフォについて】

モルフォは「画像処理／AI（人工知能）」の研究開発型企業です。高度な画像処理技術を組み込みソフトウェアとして、国内外のスマートフォン、半導体メーカーを中心にグローバルに展開しています。また、カメラで捉えた画像情報をエッジデバイスやクラウドで解析する、AI を駆使した画像認識技術を車載や産業 IoT 分野へ提供し、様々なイノベーションを先進のイメージング・テクノロジーで実現しています。

所在地：東京都千代田区神田錦町 2-2-1 KANDA SQUARE 11 階 WeWork 内

代表者：代表取締役社長 平賀 督基（まさき）、【博士（理学）】

設立：2004 年 5 月 26 日

資本金：1,783,958 千円（2022 年 10 月 31 日現在）

事業内容：画像処理および AI（人工知能）技術の研究・製品開発。スマートフォン・半導体・車載・産業 IoT 向けソフトウェア事業をグローバルに展開。

ホームページ：<https://www.morphoinc.com/>

Facebook：<https://www.facebook.com/morphoinc>

X（旧 Twitter）https://twitter.com/morpho_inc

【お問合せ先】

株式会社モルフォ 広報担当 大野

TEL：080-8433-3415

お問い合わせフォーム：<http://www.morphoinc.com/contact>

*モルフォ、Morpho およびモルフォロゴは株式会社モルフォの登録商標または商標です。