

2022年4月28日

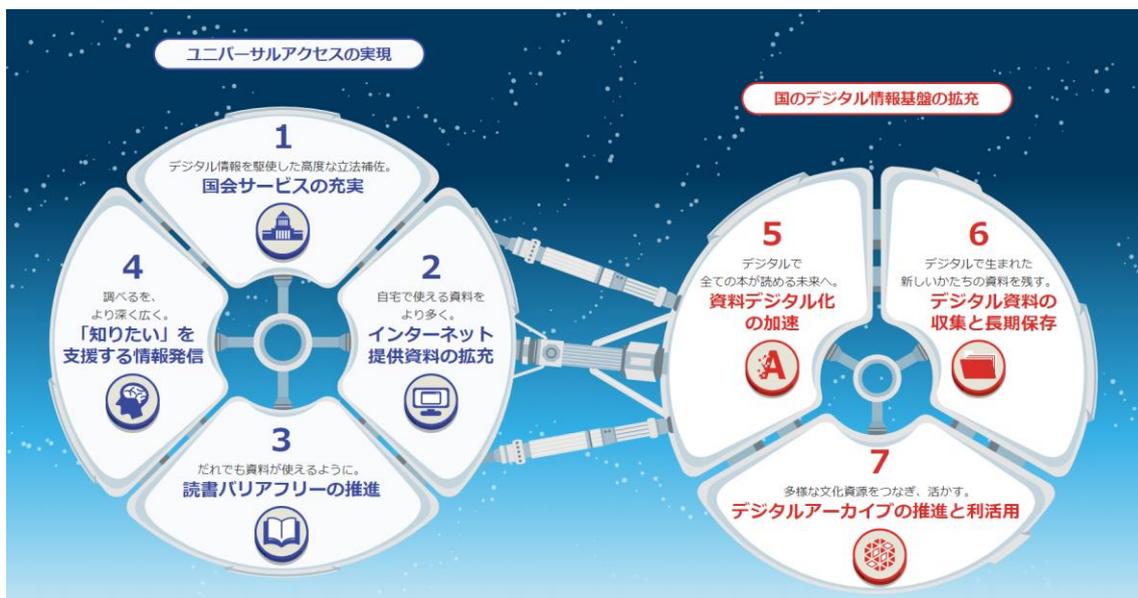
モルフォ AI ソリューションズが国立国会図書館の 最新 AI 技術を活用した OCR 処理プログラムの開発を完了

【概要】

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、国立国会図書館からの「OCR 処理プログラムの研究開発（以下 本事業）」委託事業を完了しましたので、発表します。

国立国会図書館は『ビジョン 2021-2025 国立国会図書館のデジタルシフト』の一環として、将来にわたる全ての利用者に多様な情報資源を提供するユニバーサルアクセスを実現する事業と、そのための恒久的なインフラとなる国のデジタル情報基盤の拡充を図る事業に取り組んでいます。

『ビジョン 2021-2025 国立国会図書館のデジタルシフト』：<https://vision2021.ndl.go.jp/>



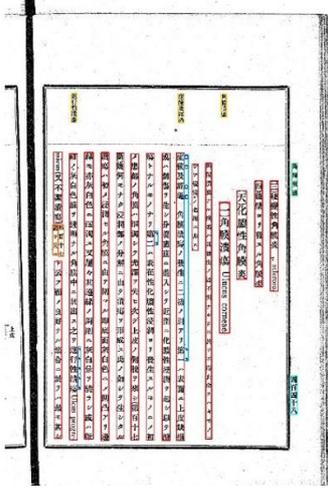
本事業では、国立国会図書館デジタルコレクション上で提供される資料画像について、今後本文テキストデータの作成を行えるようにモルフォの保有する最新 AI 技術・画像処理技術を取り入れた OCR 処理プログラムの研究開発を実施しました。また、凸版印刷株式会社（東京都文京区、代表取締役社長：磨 秀晴）の協力のもと、約 1,300 万文字の OCR 学習用データセットを構築しました。

令和 3 年度に開発した OCR 処理プログラムは多様なレイアウト・文字種に対応しており、既存の OCR サービスが対応できなかった明治期～昭和期までの複雑な資料のテキスト化が可能になりました。

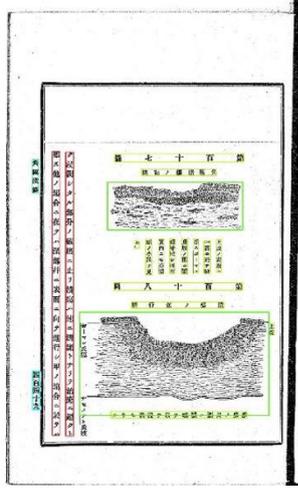
■明治期～昭和期の書籍画像（2億枚）のテキスト化処理の研究開発

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

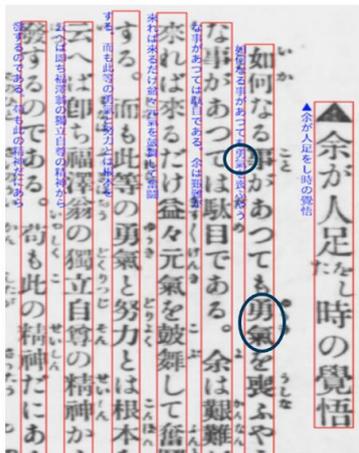


近代雑誌（大正）・昭和の雑誌レイアウトについても認識可能

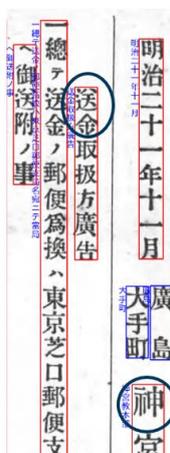


②多様な文字種（旧字旧仮名）への対応

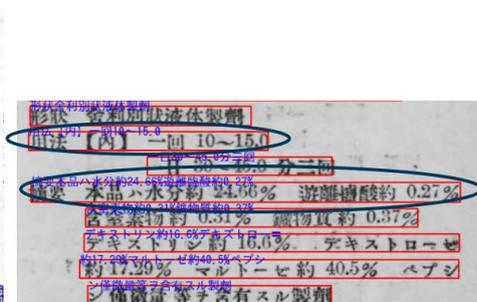
旧字体の認識



異体字→新字への変換



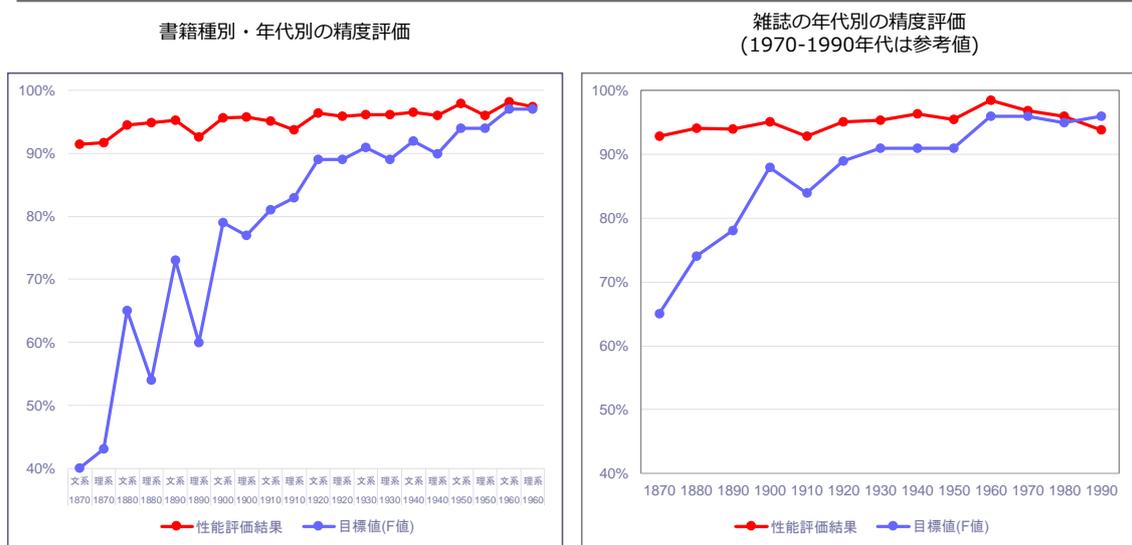
記号（〇、～、%）や英数字



③OCR 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%以上の精度で認識可能となりました。特に明治期～昭和初期の近代書籍・雑誌においては、市販 OCR の約 2 倍（約 40%→90%以上）の読み取り精度となりました。

認識精度基準と認識性能評価結果まとめ



<国立国会図書館次世代システム開発研究室様コメント>

今回の成果である日本語の OCR 処理プログラム「NDLOCR」は、NDL ラボ公式 GitHub アカウント (<https://github.com/ndl-lab>) からオープンソースで令和 4 年 4 月 25 日に公開しました。

NDLOCR は、学習用データを用意することで追加の学習が可能な OCR で、今後国立国会図書館がデジタル化する資料の全文テキストデータ作成に使用します。プログラムのほかに、開発に用いた機械学習用データセット（著作権保護期間が満了したデジタル化資料から作成した分のみ）も近々に公開予定です。日本語 OCR 全体の精度向上に資することになれば、と考えておりますので、関心のある多くの方々にご活用いただきたいと思っております。

【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 10 階

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 10 階

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（2021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>

【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。