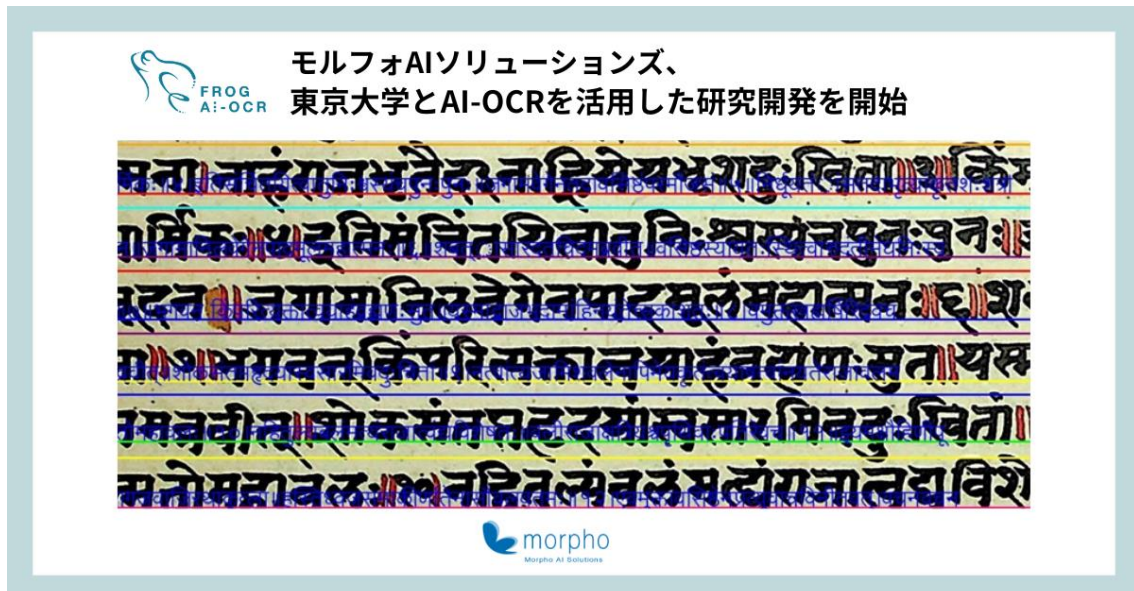


2024 年 4 月 22 日

**モルフォ AI ソリューションズ、
東京大学と AI-OCR を活用した研究開発を開始
～デーヴァナーガリー文字 OCR によってインドの文献解析に寄与～**

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、国立大学法人 東京大学（以下 東京大学）とともに、「デーヴァナーガリー文字 OCR の実用化と文献データベースの利活用にむけた応用研究」に向けて、AI-OCR を活用した研究開発を開始したことをお知らせします。



モルフォ AIS は、2022 年 7 月 1 日より大学図書館・地方図書館・自治体からのデジタル田園都市構想を背景としたデジタルアーカイブ事業や読書バリアフリー法対応のニーズを受け、市販 AI-OCR ソフトとして近代書籍対応の「FROG AI-OCR」を提供してきました。

今回、AI-OCR の開発や導入ノウハウを有するモルフォ AIS とサンスクリット文献学の専門的知識を有する東京大学、および特殊字形の AI-OCR 用学習データ制作にノウハウを有する TOPPAN 株式会社が共同で AI-OCR を開発し、サンスクリット文献に使用されているデーヴァナーガリー文字のテキストデータベース化を加速させる研究に取り組みました。

【本研究について】

本研究「デーヴァナーガリー文字 OCR の実用化と文献データベースの利活用にむけた応用研究」（JSPS 科研費基盤研究(B)）では、デーヴァナーガリー文字を読み取るための光学文字認識（OCR）ソフトウェアの開発と実用化を目指し、その OCR を用いて読み取ったサンスクリット文献群のデータベースを利活用するための応用研究を行います。デーヴァナーガリー文字とは、ヒンディー語、マラーティー語、ネパール語などの現代語のみならず、インド圏の文化や歴史などについて多くの史資料を残すサンスクリット語の表記のための主要な文字として使用されてきたインド系文字の一つです。他の人文系分野同様、サンスクリット文献学の分野においても文献のデジタルアーカイブ化やテキストデータベース化は最重要課題であり、これまでも世界各国で様々なプロジェクトが展開されてきました。

しかしながら、これらのプロジェクトはいずれも手作業によるデータ化が中心であり、個々の研究者の多大な時間と労力を要している状況です。

このような状況をふまえ、これまで手作業で行われてきたテキストデータ採取の方法を自動化するため、本研究では「FROG AI-OCR」のコアエンジンとして利用している国立国会図書館の NDLOC^(※) を改良することで、デーヴァナーガリー文字の OCR テキスト化を可能としました。

現在は、その読み取り結果を検証して OCR の精度を改善する研究段階となりますが、最終的には手作業での転記や修正作業の手間を大幅に削減することを目指しています。

今後もモルフォ AIS は、産官学との連携により、多様な用途の AI-OCR を新たに開発していく予定です。

【東京大学 加藤 隆宏 様コメント】

本研究によって開発されるデーヴァナーガリー文字 OCR は、私たちが専門的に研究するサンスクリット語写本を自動でテキストデータ化することを第一の目的としております。しかし、OCR が実用化されるようになれば、その延長線上に様々な可能性が開けていきます。

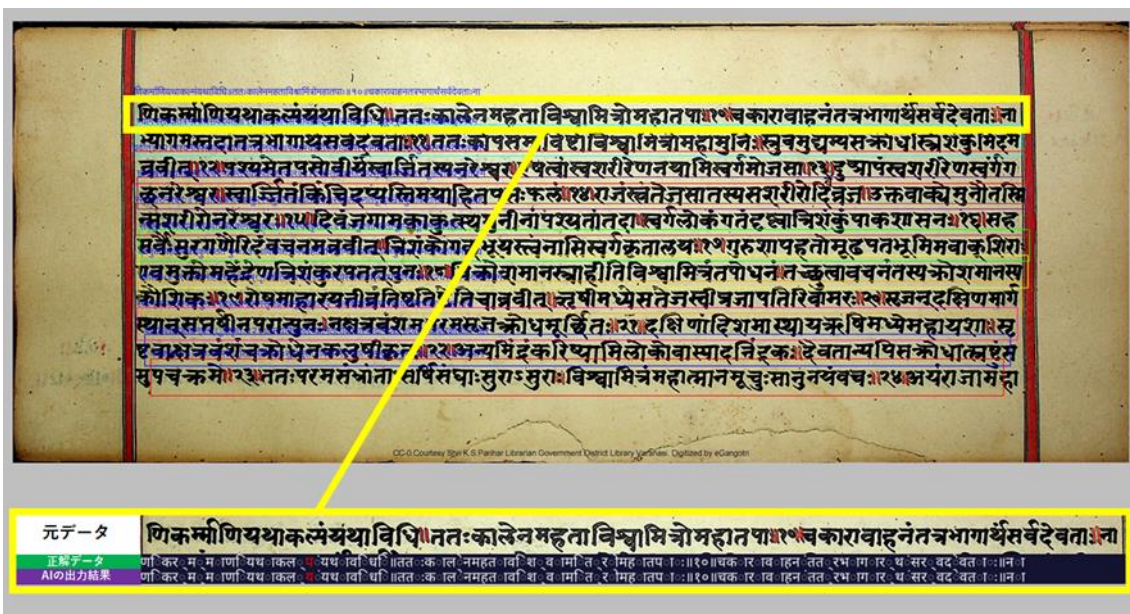
インド国内外には、いまだ多くの文献資料が手書き写本として残されており、最近では、国家的プロジェクト主導のもと、デジタル撮影・デジタルスキャンなどによって電子アーカイブ化が進められています。ただ、これだけではまだ充分とはいえず、こうした写本資料のテキストデータ化、さらにはデータの構造化が近い将来には必ず必要になってきます。OCR によるサンスクリット語写本資料の読み取りについては、これまでほとんど手が付けられてこなかった研究分野で、今回の取り組みは世界的にみても先駆的な研究プロジェクトとなっていくことが期待されます。

【モルフォ AIS 代表取締役 神田 武コメント】

デーヴァナーガリー文字 OCR プロジェクトは、技術革新を通じて文化的遺産を保存し、広く共有するという私たちのミッションに合致するものです。東京大学様との共同研究を通じて、私たちは知識の壁を取り払い、全人類が共有できる価値ある資源を創り出すことに尽力します。本研究の成功は、当社の AI-OCR 技術の応用範囲を広げ、未来の文献解析に対する大きな前進に寄与すると確信しています。貴重な機会をいただいた東京大学様、凸版印刷様に深い感謝を申し上げます。

【OCR 処理サンプル】

国立国会図書館の NDOCR を改良し、デーヴァナーガリー文字向けに追加学習をすることで、以下のサンプル画像のような処理をすることが可能となりました。



【関連製品】

・FROG AI-OCR 製品ページ

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

・FROG AI-OCR 紹介ムービー



・FROG AI-OCR 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。

【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.morphoai.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.morphoai.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 12 月 05 日

近代書籍対応 AI-OCR ソフト「FROG AI-OCR」の研究者向けパッケージプランを発表

https://www.morphoinc.com/news/20231205-jpr-mais_frog_aiocr

【注釈】

※「FROG AI-OCR」は、国立国会図書館の NDLOCR をコアエンジンとして利用しています。

※NDLOCR は、国立国会図書館が CC-BY-4.0 のライセンスで公開しているものです。

(https://github.com/ndl-lab/ndloclr_cli)

【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 や AI カメラ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（2023 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 : https://twitter.com/morpho_inc

【お問合せ先】

モルフォ AI ソリューションズ 石崎

メール : contact@morphoai.com

*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。