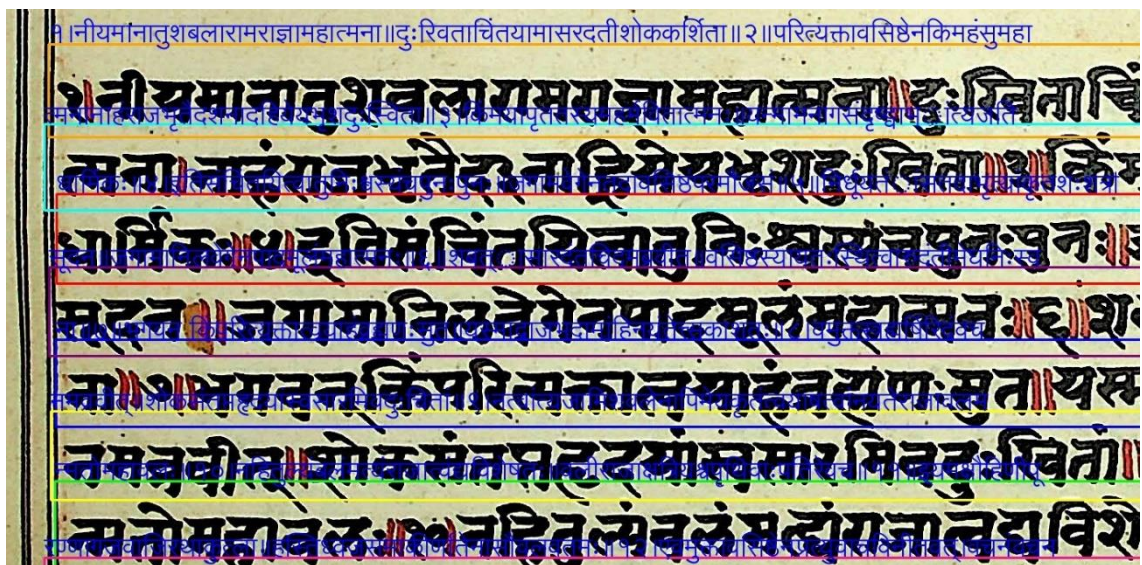


## モルフォ AI ソリューションズ、東京大学と AI-OCR を活用した研究開発を開始 ～デーヴァナーガリー文字 OCR によってインドの文献解析に寄与～

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）の「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 の提供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」の目的は、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うことです。

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と引き換えに築かれてきたものでした。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の OCR を開発し、サンスクリット文献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。この研究ではサンスクリット文献学の専門的知識を有する東京大学と AI・画像処理技術の開発者であるモルフォ AIS が共同して AI-OCR を開発することになりました。

モルフォ AIS は、複雑なレイアウトや旧字旧仮名遣いに対応可能な「FROG AI-OCR」を保有しており、ベースとなる AI-OCR を改良することで、デーヴァナーガリー文字の OCR 処理を可能としました。これにより、人手での転記作業や、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。

現在で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にあります。

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 を新規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、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に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、TOPPAN 株式会社様に深い感謝を申し上げます。

## ■ 処理サンプル

弊社保有の AI-OCR をベースに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様な処理をすることが可能となりました。

元データ	णिकर्मणियथाकस्म्यंथाविधि॥ततःकालेनमहताविश्वामित्रोमहातपाः॥१०॥चकारवाहनं तत्रभागार्थं सर्वदेवताः॥ना
正解データ	णिकर्मणियथाकस्म्यंथाविधि॥ततःकालेनमहताविश्वामित्रोमहातपाः॥१०॥चकारवाहनं तत्रभागार्थं सर्वदेवताः॥ना
AIの出力結果	णिकर्मणियथाकस्म्यंथाविधि॥ततःकालेनमहताविश्वामित्रोमहातपाः॥१०॥चकारवाहनं तत्रभागार्थं सर्वदेवताः॥ना

## ■ FROG AI-OCR 製品ページ

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

## ■ FROG AI-OCR 紹介ムービー

## OCR～校正・出力までのE2Eのソリューション



### ■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.morphoinc.com/news/20220428-jpr-mais\\_ndl](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](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](https://www.morphoinc.com/news/20220614-jpr-mais_frog_aiocr)

※「FROG AI-OCR」は、国立国会図書館の NDLOCR ([https://github.com/ndl-lab/ndlocr\\_cli](https://github.com/ndl-lab/ndlocr_cli)) をコアエンジンとして利用しています。

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

#### 【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>

X（旧 Twitter）：<https://twitter.com/MorphoAIS>

※本リリースに関するお問い合わせについては、石崎 [contact@morphoai.com](mailto:contact@morphoai.com) まで