

Press Release

報道関係各位



2022年10月11日
株式会社豆蔵

豆蔵、AI/機械学習を使用した「帳票データの構造化技術」に関する特許を取得

株式会社豆蔵デジタルホールディングスの事業会社である、株式会社豆蔵(本社：東京都新宿区、代表取締役社長：中原徹也、以下 豆蔵)は、自社で提供する「AI・RPA・DX 関連サービス」において、AI/機械学習を使用した「帳票データの構造化技術」に関する特許を取得したことをお知らせします。

■特許登録情報

発行国 : 日本国特許庁(JP)

公報種別 : 特許公報(B2)

特許番号 : 特許第 7122896 号

発明の名称 : 帳票情報処理装置、帳票情報構造化処理方法及び帳票情報構造化処理プログラム

特許権者 : 株式会社豆蔵

特許登録日 : 2022年8月12日

J-PlatPat : <https://www.j-platpat.inpit.go.jp/c1800/PU/JP-2020-013281/54ED12FC8FD49051BA962104E07E8EC3DD0B86E6DF36651D3BEFFE35CA2A7A82/11/ja>

※J-PlatPat は「独立行政法人工業所有権情報・研修館 (INPIT)」のサイトです

■特許取得の背景

電子化やデジタル化が進んできている近年においても、企業間取引における請求書・注文書・領収書などの帳票は紙媒体により交わされるケースが多く、また、各帳票は取引先企業ごとに書式や表記の形式が異なることから、多くの企業が帳票の情報をシステムに手入力しているのが現状です。これは、業務効率が悪いことに加え、入力ミスが増加する原因となります。

■特許概要(※1)

本特許技術は、上記の課題を解決すること目的として研究・開発されました。

本発明で定義される「帳票情報処理装置」は、実際に構造化データの対象となる帳票を読み取る前に、事前準備(事前学習)として管理者等が学習データ登録を実施し、利用時(帳票読み取り時)においても利用者が随時精度改善(チューニング)を実施することで、利用する回数を重ねるほど、変換精度が向上する仕組みです。

また、高価な専用スキャナ機器や OCR ソフトは不要で、市販のスキャナ機器や OCR ソフトを使用して処理を行うことが可能です。

学習データ登録および帳票読み取りの各作業工程の主なフローと、各フローの処理内容は以下のとおりです。

① 学習データ登録

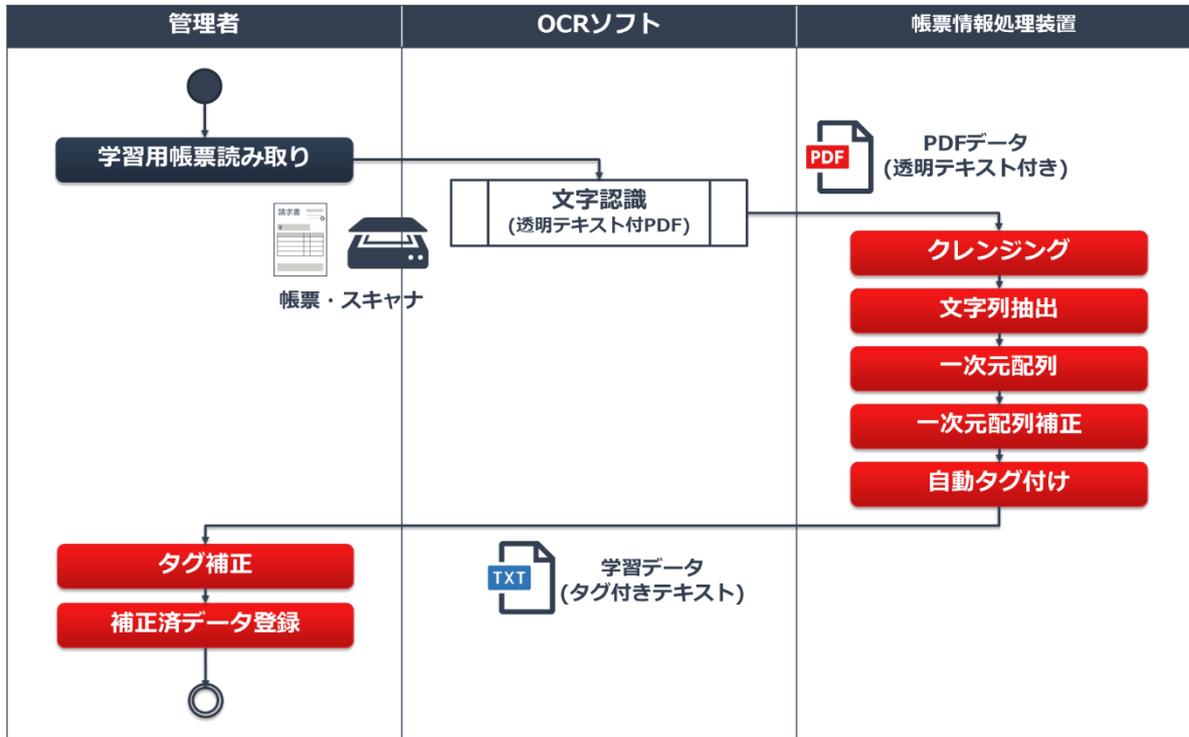


図1. 学習データ登録のフロー

学習用帳票読み取り	学習用帳票(サンプル帳票)をスキャナで読み取り OCR で文字認識(PDF 化)
クレンジング処理	PDF データ内の透明テキストデータを抽出して誤認識されたデータ等を除去
文字列抽出	特定の文字列や記号 (日付、伝票 No、〒、¥等) のパターンから対象文字列抽出
一次元配列	抽出した文字列を一次元配列へ展開し文書内の相対位置を判定
一次元配列補正	データ型を判定して相対位置を補正
自動タグ付け	各文字列の意味解釈を行って文字列にタグを付与して学習データとして保存
タグ補正	管理者が自動タグ付けによる誤りを補正
補正済データ登録	管理者が補正したタグ付けデータ(学習データ)を保存

② 帳票読み取り

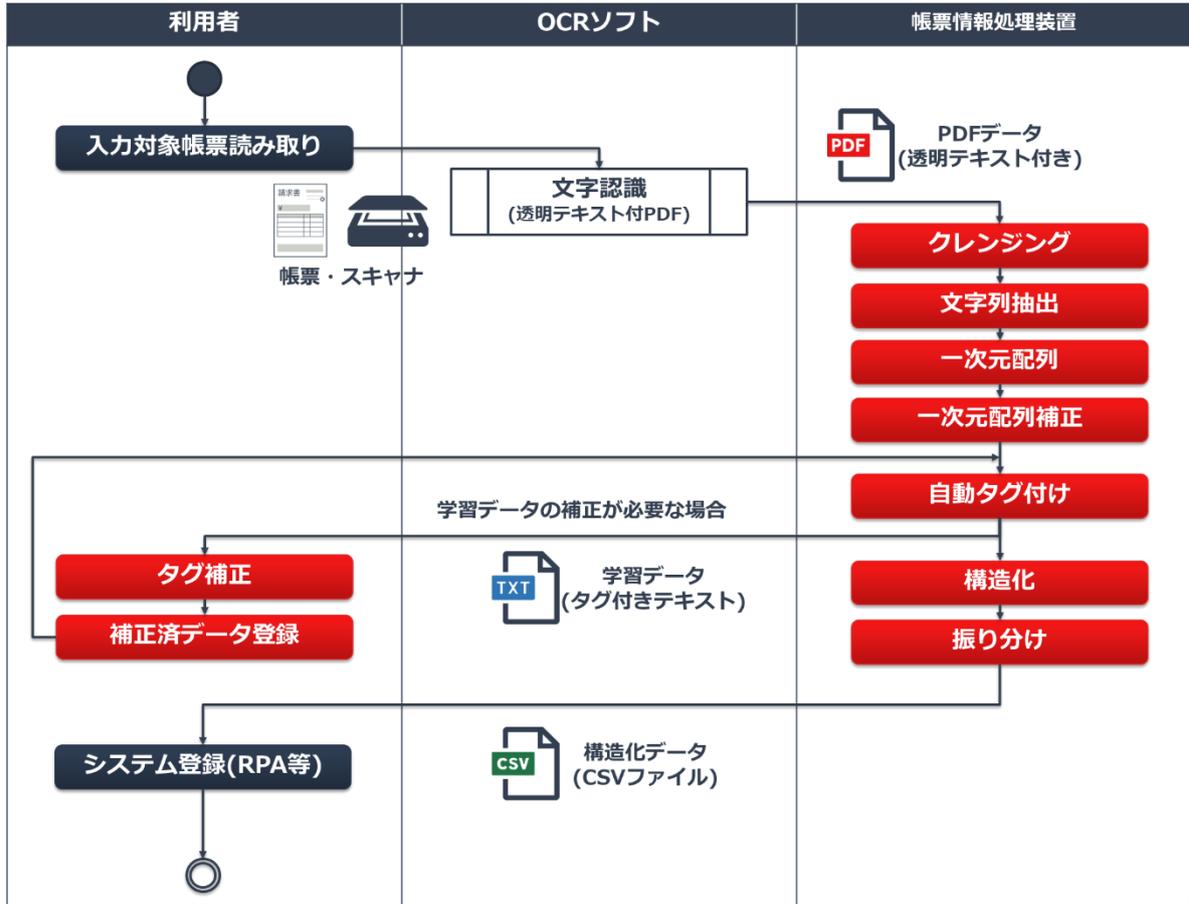


図 2. 帳票読み取りのフロー

入力対象帳票読み取り	システムへ登録するデータを含む帳票をスキャナで読み取り OCR で文字認識(PDF 化)
クレンジング処理	学習データ登録と同様
文字列抽出	学習データ登録と同様
一次元配列	学習データ登録と同様
一次元配列補正	学習データ登録と同様
自動タグ付け	学習データ登録と同様
タグ補正	利用者が自動タグ付けによる誤りを補正
補正済データ登録	利用者が補正したタグ付けデータ(学習データ)を保存
構造化	入力対象帳票の内容を構造化データ(CSV データ等)へ変換する
振り分け	帳票種別(請求書、注文書等、見積書)や取引先毎にデータを振り分ける
システム登録(RPA 等)	構造化データされたデータを RPA 等でシステムへ自動登録を行う

※1：本概要説明は特許出願文を要約した内容であり、出願内容の全内容を網羅し、解釈を完全に保証するものではありません。より精緻な情報は、先述の特許情報システム(J-PlatPat)のリンク先にある公開文献をご参照下さい。

■特許技術によって得られる効果

本特許技術は、これまで広く利用されている紙帳票を電子化するための OCR 処理に対して、AI/機械学習を用いたデータ補正処理を拡張することにより、高い精度で構造化データ(CSV ファイル等)へ自動変換を可能としました。これにより、RPA ロボット等を活用したシステムへの自動登録が可能となり、システムへのデータ登録の負担軽減とデータの品質向上促進を期待できるものと考えております。

■今後の計画

当社の AI・RPA・DX 関連サービスにおける RPA 導入・運用支援『RPAOffice』における活用や、対話型 AI エンジン『MZbot』における機能実装を計画しております。また、AI・RPA・DX 関連サービスに関連する特許出願は本件以外にも 1 件(特開 2022-019441)ございますが、今後も当社コンサルティングサービスや製品開発の過程で得られた新規性が認められるアルゴリズムや機能については随時特許出願の検討を進めてまいります。

■ RPA 導入・運用支援『RPAOffice』

RPA 導入・運用支援『RPAOffice』の詳細については当社公式サイトをご覧ください。

<https://www.mamezou.com/services/strategic/rpaoffice>

■ MZbot 製品サイト

MZbot に関する詳細については製品サイトをご覧ください。

<https://www.mzbot.jp>

■ 商標・特許について

- MZbot および RPAOffice は株式会社豆蔵が日本における登録商標です。
- その他記載されているサービス名、製品名などの固有名詞は、各社の商標または登録商標です。

【株式会社豆蔵 概要】

所在地 : 東京都新宿区西新宿2-1-1 新宿三井ビルディング34階

設立 : 2006年10月 (2000年5月より事業開始)

代表者 : 代表取締役社長 中原 徹也

資本金 : 3億1,000万円

株主 : 株式会社豆蔵デジタルホールディングス 100%

URL : <https://www.mamezou.com/>

<導入相談に関するお問い合わせ>

株式会社豆蔵 営業本部 第四営業部

Tel : 03-5339-2114

WEB : <https://www.mzbot.jp/contact>

<報道関係者からのお問い合わせ>

株式会社豆蔵 管理本部管理部 広報宣伝グループ

Tel : 03-5339-2114

WEB : <https://www.mamezou.com/contact>