

安定かつ高精度な文章抽出で ユーザーの翻訳業務をサポート



「文章の抽出精度が飛躍的に向上し、問い合わせも20%減。ユーザーからの評価も高く、要望に著実に応えられつつあると実感しています」

テクニカルサポート リム エイデン氏

導入製品

- Adobe Acrobat Service API

導入メリット

文章抽出精度の高さ

PDFのレイアウトを維持したまま
文章の抽出が可能

厳格なセキュリティ要件を 満たす

社外秘の文書を多く扱う企業でも
安心して使えるAPIセキュリティ

安定稼働が実現

アクセスが集中する時間も
処理速度を落とさず快適に使える

多種多様なファイルの 翻訳対応を実現

製造業を中心とするユーザーから
要望の多いOCRにも対応可能

八楽株式会社

2009年設立。自動翻訳とポストエディット（後編集）で翻訳の品質やスピードを向上させるAI自動翻訳ツール『YarakuZen』を開発・提供している。翻訳機能のみならず、共有機能や外注業者への依頼機能も実装。『YarakuZen』及び機械翻訳エンジン開発で培ったノウハウ・リソースを活用し、AI関連の研究も行っている。

八楽株式会社が開発・提供を手がける『YarakuZen』は、PDFをはじめとする9種類のフォーマットや、35の言語に対応できるAI自動翻訳プラットフォームだ。同社は『YarakuZen』で翻訳するPDFの文章を抽出してWordに変換するAPIにAdobe Acrobat Service APIを採用。高い抽出精度により翻訳の精度も向上し、問い合わせ件数も2割程度減少するといった効果が表れている。

■ 導入の経緯

PDFからWordに変換する工程で発生する不具合を解消するため、複数のAPIを検討

近年は海外に拠点を置くグローバル企業が増加し、翻訳サービスの需要も増加傾向にある。その中であらゆる業界の翻訳ニーズに対応するのが、八楽株式会社（以下、八楽）のAI翻訳プラットフォーム『YarakuZen』だ。

「AI翻訳+あなた（ユーザー）」で成長する『YarakuZen』は、会社や各部門で頻出する専門用語をAIに学習させることで翻訳を最適化する仕様や、二次利用のリスクなく翻訳できるセキュリティ面などにより、幅広い業種・業態のユーザーから高い評価を得ている。カスタマーサポートマネージャーの武田 啓子氏は、『YarakuZen』についてこう説明する。

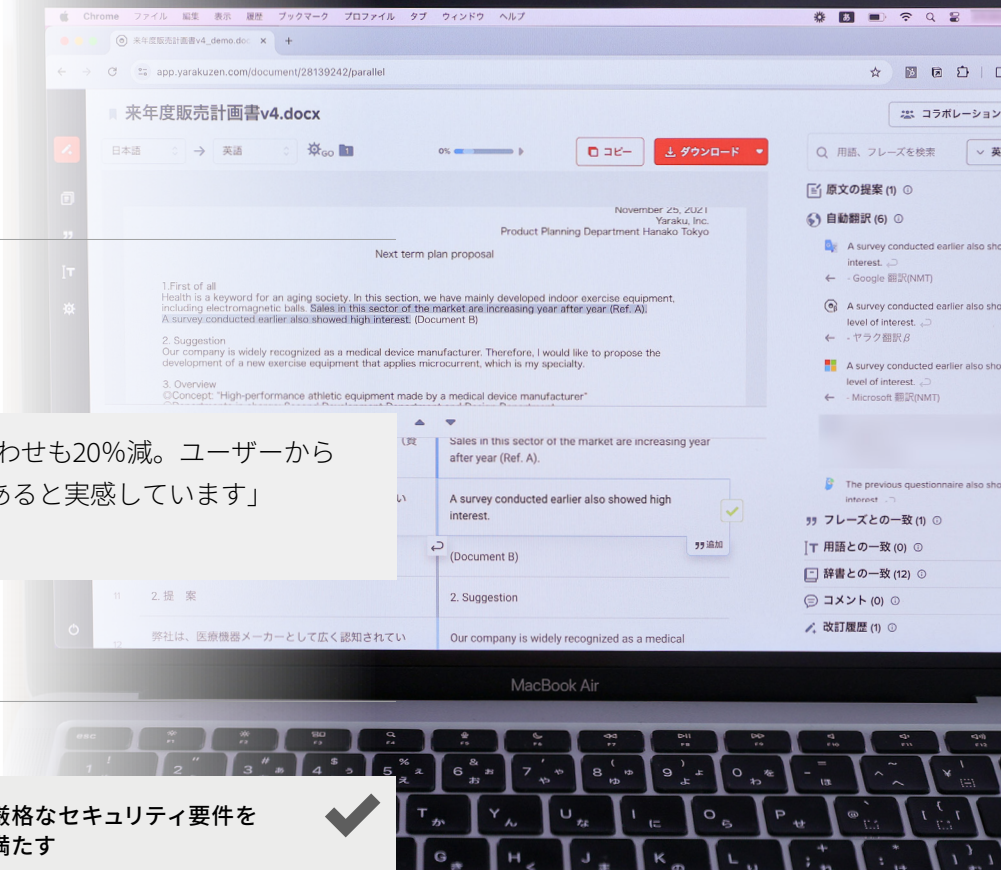
「ユーザー層は自動車や電気機械メーカーなど、製造業のお客が多くいらっしゃいますが、業種・業態の幅はとても広いです。営業のプレゼン資料、論文、設計図といった書類の翻訳はもちろん、社内外へのメール文章の翻訳も気軽に行えるため、さまざまなビジネスシーンでご活用いただけます」

中でも特に需要の多いフォーマットの1つであるPDFを翻訳する場合、PDF内の文章はWordに変換される仕様となっている。しかし以前使用していたAPIでは、文章の抽出精度が不十分で、ユーザーから問い合わせが来ることも多かった。テクニカルサポートのリム・エイデン氏は、当時の状況を次のように振り返る。

「数年前に使用していたAPIでは、文章として正しく抽出できる範囲が狭く、抽出できても文字化けした状態でWordに変換されてしまうことがしばしばありました。また、ファイルを取り込んでも翻訳されずエラーになる確率が高かったことにも頭を悩ませていました」



テクニカルサポート
リム エイデン氏





カスタマーサポートマネージャー
武田 啓子氏

USER PROFILE

八楽株式会社

<https://www.yaraku.com/ja/>
<https://www.yarakuzen.com/>
所在地：東京都渋谷区千駄ヶ谷5-27-5 リンクススクエア新宿16階
創業：2009年8月
資本金：3億6396万円（資本準備金含む）
従業員数：23名（〇〇年現在）
事業内容：多言語コミュニケーションツールの企画・開発・運用



詳細情報

<https://developer.adobe.com/document-services/apis/pdf-services/>



アドビ株式会社
〒141-0032 東京都品川区大崎1-11-2
ゲートシティ大崎イーストタワー
www.adobe.com/jp/
Adobe
345 Park Avenue
San Jose, CA 95110-2704
USA
www.adobe.com

その後、Adobe Acrobat Service APIを含むAPIを複数比較。Adobe Acrobat Service APIを使用した文章の抽出精度が圧倒的に高いと判断した同社は、Amazon Web Serviceを通じて利用を開始。さらに2023年12月にはアドビの日本法人と直接契約することになった。

■ 選択のポイント
翻訳の必須条件となる文章の抽出精度の高さが決め手に

八楽がAdobe Acrobat Service APIを利用し続けているポイントは、主に4つある。文章の抽出精度が高い点、セキュリティ面でも信頼できる点、クラウドベースで提供されている点、そして画像データ内の文字を認識し、テキストデータへ変換するOCRにも対応している点だ。

「抽出精度の高さについては、変換したWordを開いた時点で他のAPIとの違いがすぐに分かりました。日本語から英語に翻訳すると、英文の方がどうしても一文が長くなるのでレイアウトが崩れやすくなりますが、Adobe Acrobat Service APIで抽出した文章はレイアウトの崩れがほぼありませんでした」（エイデン氏）

2023年には、政府のセキュリティ要件を満たすクラウドサービスを評価するISMAPにAdobe Document Cloudが登録された。アドビのこうした高いセキュリティ水準やネームバリューからの安心感が、翻訳文章の漏洩リスクを懸念するユーザーの信頼獲得に繋がられているといわれ、同社でもそれを実感している。

クラウドベースであるメリットに関しては、同社の都合もあった。オンプレの場合は自社に別途サーバーに入れて運用しなければならない上に、日々のメンテナンスに対応する分の人件費や災害などの予期せぬ事故が発生することが懸念点として挙げられる。

「サーバーがダウンして『YarakuZen』が動かなくなるような自体は何としても避けなければいけません。そうしたトラブル防止には日ごろのメンテナンスを維持する必要がありますが、クラウドのAdobe Acrobat Service APIであれば、そのような負担をかけずに運用ができます」（エイデン氏）

OCR機能については以前よりユーザーからのニーズが多く、多種多様なレイアウトで構成されたファイルの翻訳対応を実現するためには、OCRでテキストを認識できるAPIが必須条件だったという。

■ 導入効果
テクニカルサポートへの問い合わせが20%減少

『YarakuZen』で翻訳を行うには、PDFから正確に文章を抽出できることが必要不可欠である。その点で、抽出精度の高さが要求される『YarakuZen』に見合うAPIが、Adobe Acrobat Service APIだった。導入後、文章の抽出精度は飛躍的に向上。PDFのレイアウトを崩さず、非常に安定した精度で読み取りが可能になった。武田氏、エイデン氏の両者とも、その導入効果を実感している。

「長年『YarakuZen』を愛用いただいているユーザーから、抽出精度の高さをご評価いただける機会が増えました。それを象徴するかのように、問い合わせの数は体感で20%程減少しています」（エイデン氏）

また、かつて自社のサーバーにPDFコンバーターAPIを入れて運用していた頃は、プラットフォームへのアクセスが集中すると急激に処理速度が遅くなってしまいう事も多かったという。しかしクラウドベースに切り替えた現在は、ユーザーからのアクセスが集中するピークタイム（午後2～4時頃）においても、処理速度が落ちることはほぼなくなった。そのため、安定的に質の高いサービスの提供が可能となっている。

■ 今後の展望
より安定した文章の抽出や、対応できるフォーマットの増加も検討

近年盛んにDX化が叫ばれているためか、ユーザーから特に寄せられるのはOCRに関する要望だ。複雑な構造を持つPDFの抽出精度もさらに安定させていき、ゆくゆくは「スキャンした手書きの書類を『YarakuZen』で翻訳したい」といった声にも応えることを検討しているという。

ほかにも、『YarakuZen』で翻訳するためにアップロードしたPDFは現在Word形式でのみユーザーがダウンロードできる仕様になっているが、今後はさまざまな利用シーンを想定してPowerPointに変換できるオプションの開発も視野に入れている。

最後に、今後のアドビへの要望も伺った。

「以前と比較して、Adobe Acrobat Service APIを利用し始めてからはユーザーの要望に著実に応えられつつあると実感しています。設計図のような図表や画像が多い複雑なPDF、文章が画像内に組み込まれているPDFについても、より確実にテキストデータとして抽出できるよう、アドビさんと協力しながらサービスを改善していきたいですね」（エイデン氏）

※掲載された情報は、2024年2月現在のものです。