ドキュメント翻訳

# 概要

ドキュメント翻訳は [Azure Translator サービス](https://azure.microsoft.com/en-us/services/cognitive-services/translator/)の新機能であり、大量の複雑なドキュメントを必要とする企業、翻訳会社、および消費者が、元のドキュメントの構造と形式を保持したまま 1 つ以上の言語に翻訳できます。 **Translator サービスでサポートされている** 111 の言語と方言**のいずれかで**、テキスト、HTML、Word、Excel、PowerPoint、Outlook、PDF、Markdown など、さまざまなファイル形式のドキュメント全体を非同期に翻訳します。

市場に出回っている標準的な翻訳サービスでは、プレーンテキストまたは html のみを受け入れ、リクエストの文字数に制限を設けています。大きなドキュメントを翻訳するユーザーは、ドキュメントを解析してテキストを抽出し、それらを小さなセクションに分割して個別に翻訳する必要があります。文が不自然なブレークポイントで分割されると、コンテキストが失われ、最適でない翻訳になる可能性があります。翻訳結果を受け取ったら、お客様は翻訳された部分を翻訳されたドキュメントに統合する必要があります。これには、どの翻訳された作品が元のドキュメントの同等のセクションに対応するかを追跡することが含まれます。顧客がリッチコンテンツを含む複雑なドキュメントを翻訳したい場合、問題は複雑になります。



ドキュメント翻訳は、お客様が簡単に翻訳できるようにします。

1. 大量の大きなドキュメント、
2. さまざまなファイル形式のドキュメント、
3. 元のレイアウトと形式を保持する必要があるドキュメント
4. ドキュメントを複数のターゲット言語に翻訳します。

# ユーザーエクスペリエンス

ユーザーは、ソースドキュメントとターゲットドキュメントの場所、およびターゲット言語のリストを指定して、ドキュメント翻訳サービスにリクエストを行います。このサービスは、ユーザーが翻訳のステータスを追跡できるようにする識別子を返します。ドキュメント翻訳は、ソースの場所から各ドキュメントを非同期的に取得し、ドキュメント形式を認識し、適切な解析手法を適用してドキュメント内のテキストコンテンツを抽出し、テキストコンテンツをターゲット言語に翻訳します。次に、翻訳されたドキュメントを再構築し、ソース ドキュメントに存在するレイアウトと形式を保持し、翻訳されたドキュメントを指定された場所に保存します。ドキュメント翻訳は、ジョブレベルまたはドキュメントレベルで翻訳のステータスを更新します。

ユーザーは、カスタム翻訳ポータル、カスタム用語集、またはその両方を使用して作成されたカスタム モデル ID を、ドキュメントの翻訳要求の一部として提供できます。ドキュメント翻訳は、特定の用語を保持し、翻訳されたドキュメントにドメイン固有の翻訳を提供して、このようなカスタマイズを適用します。