文件翻译

# 概述

文档翻译是 [Azure 翻译服务](https://azure.microsoft.com/en-us/services/cognitive-services/translator/)中的一项新功能，它使需要将大量复杂文档翻译成一种或多种语言的企业、翻译机构和使用者能够保留原始文档的结构和格式。它以各种文件格式异步翻译整个文档，包括文本**、HTML、Word、Excel、PowerPoint、Outlook、PDF 和 Markdown**，翻译服务支持**的 111 种语言和方言**中的任何一种。

市场上的标准翻译产品只接受纯文本或 html，并限制请求中的字符数。翻译大型文档的用户必须解析文档以提取文本，将它们拆分为更小的部分并单独翻译。如果句子在不自然的断点中拆分，则可能会丢失上下文，从而导致次优翻译。收到翻译结果后，客户必须将翻译后的部分合并到翻译后的文档中。这涉及跟踪哪些翻译片段与原始文档中的等效部分相对应。当客户想要翻译具有丰富内容的复杂文档时，问题就会变得复杂。



文档翻译使客户能够轻松翻译：

1. 大量大型文档，
2. 各种文件格式的文档，
3. 需要保留原始布局和格式的文件，以及
4. 文档转换为多种目标语言。

# 用户体验

用户向文档翻译服务发出请求，指定源文档和目标文档的位置以及目标语言列表。该服务返回一个标识符，使用户能够跟踪翻译的状态。异步，文档翻译从源位置拉取每个文档，识别文档格式，应用正确的解析技术提取文档中的文本内容，将文本内容翻译成目标语言。然后，它重建翻译后的文档，保留源文档中存在的布局和格式，并将翻译后的文档存储在指定位置。文档翻译在作业或文档级别更新翻译状态。

用户可以提供使用自定义翻译门户和/或自定义词汇表构建的自定义模型 ID，作为翻译文档请求的一部分。文档翻译应用此类自定义，保留特定术语，并在翻译后的文档中提供特定领域的翻译。