

[Imprimir esta página](#)

TRADUCIR UN PDF SIN MORIR EN EL INTENTO

Publicado el 19 Abril 2016

Publicado en: [Artículos y tutoriales](#)[Marco Cevoli](#)

Visualizaciones: 7850

(2 Votos)Etiquetado como :

[Adobe Reader](#)[Acrobat](#)[Ocr](#)[Microsoft](#)

El formato PDF es uno de los enemigos más temidos por los profesionales de la traducción. Para utilizar una herramienta de traducción asistida por ordenador (TAO) en la traducción o análisis de un PDF, es necesario convertirlo a un formato editable. Esta conversión puede ser más o menos difícil (llegando a imposible en ocasiones) en función del tipo de PDF que se recibe. Distinguir a primera vista los diferentes tipos de PDF es esencial para la aplicación de un procedimiento de conversión correcto y el uso de una herramienta adecuada: así se consigue reducir al mínimo el tiempo que transcurre entre la recepción del archivo y el inicio real de la traducción.

PDF son las siglas de Portable Document Format, un formato de archivo desarrollado por Adobe en 1993 para representar documentos independientemente del hardware y el software utilizado para generarlos o visualizarlos. En definitiva, los PDF se visualizan y procesan del mismo modo indistintamente del equipo que se utilice para ello. Esta característica ha hecho del formato PDF uno de los métodos preferidos para compartir documentos. Para muchas personas, crear un PDF de un documento se ha convertido en el equivalente de «hacer una fotocopia digital» del mismo, con todas las ventajas que ello supone en términos de practicidad, pero sin tener en cuenta los inconvenientes que se presentan a la hora de editar o traducir el documento PDF en sí.

Cuando recibe un archivo PDF se recomienda comprobar la naturaleza del contenido. Un PDF suele contener elementos de distintos tipos. Algunos elementos no pertenecen al texto visible: las llamadas «propiedades», como pueden ser el nombre del autor, título, etc. El resto de elementos forman el documento y, en general, son: texto, imágenes de mapa de bits (fotografías) y gráficos vectoriales (líneas, algunos tipos de diagramas). Para nuestros propósitos, es importante comprobar si el texto del documento aparece como tal o puede seleccionarse. Para saberlo, basta con abrir el documento utilizando Adobe Reader (o [cualquier otro lector de PDF](#)) y hacer clic en el icono de selección de texto situado en la barra de herramientas o aumentar el zoom del documento. Si en algún momento el texto se vuelve borroso y pixelado, significa que estamos frente a un **escaneado**. Si, por el contrario, se puede

seleccionar texto o hacer zoom sin que el texto pierda resolución, significa que el PDF se ha **generado con una aplicación**. Algunos programas de escaneado son capaces de producir documentos PDF que preservan la apariencia de la página escaneada, pero que, al mismo tiempo, reconocen el texto. En estos casos, el texto es seleccionable, aunque la página se visualice como una copia exacta del documento de papel del que deriva. En estas ocasiones es recomendable comprobar la ortografía y la exactitud del texto, ya que el OCR tiene cierto margen de error.

Para saber qué aplicación genera el PDF tiene que pulsar CTRL+D (o Archivo | Propiedades del documento) y leer la información contenida en la ficha Descripción. En el apartado «Aplicación de» aparece el nombre del programa utilizado para crear el documento PDF. Cuando esto sucede, lo ideal es pedirle al cliente que nos envíe el archivo editable, indicando que se está seguro de su existencia (se puede leer en las propiedades del mismo documento). Por lo general, para convencer al cliente basta con decir que el no hacerlo le supondrá un gasto añadido para cubrir los costes del proceso de conversión. Evidentemente, este tipo de negociación depende de la relación que se establece con el cliente y la potencia comercial en cada caso concreto. Para ser honestos, es muy probable que, especialmente en el caso de las organizaciones multinacionales, aquellos que envían el archivo PDF no tengan a su disposición el archivo editable. A menudo, los servicios de DTP se realizan en la oficina central y las sucursales sólo reciben el PDF final para imprimir a nivel local. La traducción puede ser algo que se ha requerido en un segundo momento y entonces buscar la fuente original puede ser muy laborioso.

Si, a pesar de estos esfuerzos, no está disponible el archivo original, hay varias opciones para exportar el texto. Resulta interesante señalar que ninguna de las opciones dará como resultado un archivo completamente idéntico al original, sobre todo si contiene mapas de bits y cierto grado de formato, sin entrar en la fuente utilizada en el texto. El método elegido y, por tanto, el grado de precisión, dependen también de la finalidad de la extracción del texto. Hay dos posibilidades:

1. tener el texto a disposición con el fin de realizar un recuento o análisis;
2. crear un archivo editable lo más similar posible a la original.

En el primer caso, ni siquiera resulta necesario recurrir a la extracción. Si el texto del PDF está codificado como texto (tal como veíamos antes), se pueden utilizar las siguientes herramientas:

- PractiCount and Invoice (<http://www.practiline.com/>)
- Total Assistant (<http://www.surefiresoftware.com/totalassistant/features.php>)
- AnyCount (<http://translation3000.com/products/anycount/>)
- Translator's Abacus (<http://www.globalrendering.com/download.html>)
- QuickCount (<http://www.verbumsoft.com/products.aspx?p=3>)
- TIFF PDF Counter(<http://www.tiffsoftware.com/tiff-PDF-counter.html>)

Si el archivo en cuestión no ocupa más de 1 MB, no resulta necesario instalar ningún programa, ya que existen herramientas gratuitas en línea que permiten contar el número de palabras que contiene un PDF sin necesidad de copiar y pegar el texto:

- Word Counter Tool (<http://wordcounttools.com/>)

Si, por el contrario, no se desea utilizar el software mencionado y se dispone de Adobe Acrobat (no Adobe Reader), se puede extraer el texto de la siguiente manera:

- abrir el archivo PDF con Adobe Acrobat
- en el menú Archivo, guardar el documento como RTF o DOC

Dependiendo del tipo de documento, puede ser necesario aplicar una o más macros para fijar el formato. Por ejemplo, **esta macro de Word** restaura los retornos de carro (el enlace es a una copia del sitio archivado en www.archive.org debido a que www.terminologymatters.com no está disponible). Otra macro muy eficaz, en este caso para OpenOffice, es **PerfectePub**.

Si no se dispone de Adobe Acrobat:

- abrir el archivo con Adobe Reader;

- elegir la herramienta selecciona texto;
- seleccionar todo el texto (posicionar el cursor en la primera palabra del documento, ir al final y hacer clic en la última palabra mientras se pulsa la tecla Shift);
- copiar el texto (CTRL+C);
- abrir Word o cualquier otro editor de texto y
- pegar el texto (Ctrl+V).

Evidentemente, esta opción también se puede emplear en el primer caso cuando el texto para traducir o analizar solo forma una parte de todo el documento.

Para mantener el formato existen dos opciones: Se puede utilizar uno de los muchos programas que convierten PDF a Word o directamente utilizar un programa de reconocimiento óptico de caracteres (OCR) tipo FineReader, OmniPage, ReadIris, etc. En general, no resultan muy aconsejables los programas que realizan una conversión directa sin intervención manual. Estos programas suelen crear documentos de Word que mantienen visualmente el aspecto del PDF original, pero que alcanzan este resultado a través de un formateo complicadísimo, repleto de cuadros de texto, intervalos de sección, columnas, estilos, interlineados... En cuanto nos pongamos a trabajar con el documento, por ejemplo, ya sea eliminar una frase o abrirlo con un programa de traducción asistida, el formato se descompondrá y complicará el trabajo con creces. Se aconseja realizar la conversión con un programa de OCR (uno de los que mejores resultados nos dio fue Abbyy FineReader), y modificar manualmente la configuración por defecto, o, en su defecto, indicar al programa la distribución de los diversos elementos de la página. Para más informaciones acerca del manejo de FineReader consultar el artículo «[Riconoscimento ottico dei caratteri con Abbyy FineReader](#)» (de momento solo en italiano).

Si el cliente no solo necesita conservar el formato, sino también reconstruir el archivo completamente (siempre hablando del caso de no disponer del archivo a partir del cual se ha creado el PDF), tenemos dos posibilidades:

1. trabajar con un programa de publicación de escritorio (DTP) tipo QuarkXPress, InDesign, etc., utilizando el PDF original como modelo, por ejemplo en una capa de fondo, o
2. utilizar Infix, un editor de PDF distribuido por Icení.

La versión «Professional» de [Infix](#) (que cuesta alrededor de 150 €), contiene una función útil para exportar el texto de un PDF a formato XML. Este archivo XML puede traducirse con una herramienta CAT (por ejemplo, OmegaT, que desde la versión 2.3.0 tiene un filtro para traducir directamente este formato de archivos. Se puede consultar el [tutorial detallado](#) en el sitio de OmegaT, en inglés). Una vez traducidos los archivos, hay que importarlos al PDF original, siempre utilizando Infix Professional. En el sitio de Infix se puede disfrutar de un [vídeo explicativo](#) del proceso completo.

Si en cambio, tenemos que trabajar con un programa de DTP (Quark, InDesign, etc.), el PDF original se usará como modelo en una capa separada. Es recomendable la lectura del siguiente artículo (en inglés) para ampliar información: <http://www.proz.com/translation-articles/articles/560/1/Translation-and-DTP-of-a-PDF-File>

Quien no quiera invertir en un programa de OCR porque considera que lo utilizará muy de tanto en tanto, puede usar uno de los distintos convertidores en línea, como Zamzar (<http://www.zamzar.com>). Lo anterior se aplica, repetimos, a un PDF generado por las aplicaciones. En el caso de que el texto contenido en el archivo PDF se componga de imágenes (caso típico de un fax que se haya digitalizado con un escáner), la única forma de exportarlo a un formato editable es utilizar un programa de OCR.

Aparte de la dificultad que entraña un archivo PDF en sí, los niveles de seguridad de este formato suponen una complicación añadida. Resulta posible establecer dos niveles de seguridad a través de una contraseña de usuario y una contraseña de permisos o maestra. La primera impide la apertura del documento, mientras que la segunda restringe una o varias operaciones, como imprimir, copiar o editar el texto; añadir notas; etc. Por lo tanto, si el autor del PDF ha optado por establecer una contraseña, será imposible realizar cualquier acción sobre el documento. En ese caso, es necesario ponerse en contacto con el cliente y solicitar la contraseña. Sin embargo, de no ser lo anterior posible, es bueno saber que hay muchas herramientas que pueden descifrar o eliminar rápidamente la contraseña de permisos. Solo hay que buscar en Google «desbloquear PDF» y aparecen infinidad de herramientas en línea, como <http://smallpdf.com/es/desbloquear-pdf>, que permite realizar la operación en modo de conexión segura y además cuenta con un convertidor de archivos PDF. Los PDF protegidos con contraseña de usuario que impide la apertura del

mismo son más complicados de desbloquear. Si la encriptación del documento no es demasiado fuerte, podemos recurrir a herramientas como la anterior. Si, por el contrario, la contraseña es segura, los programas recurren a métodos de «fuerza bruta» que pueden tardar horas o incluso días en descifrar la clave. Sin embargo, queremos indicar que el uso de estos instrumentos puede infringir los derechos de propiedad del documento y Qabiria no alienta su uso en ningún caso.



MARCO CEVOLI

[LinkedIn](#) / [Twitter](#) / [Google+](#)

Sitio Web: www.marcocevoli.com

Otros artículos de Marco Cevoli

[Herramientas gratuitas para traductores II: Notepad++](#)

[Herramientas gratuitas para traductores I: ApSIC Xbench](#)

[Una cálida bienvenida a nuestra nueva incorporación](#)

[¿Qué hace un realizador de vídeo en Qabiria?](#)

[¿Por qué tendría que traducir voluntariamente para TED?: 5 buenas razones](#)

0 Comments

Qabiria

 Login ▾

 Recommend

Sort by Best ▾



Start the discussion...

Be the first to comment.

ALSO ON QABIRIA

PayPal come metodo di gestione di incassi e pagamenti

1 comment • a month ago



elisabetta — La nostra agenzia pur essendo tra le prima in Francia non utilizza il pagamento online visto che che si possono

Condividere un progetto con OmegaT e Subversion

3 comments • a month ago



Carmelo Serraino — Ottimo lavoro, grazie per avermenlo segnalato e a presto!

 Subscribe

 Add Disqus to your site Add Disqus Add

 Privacy