


# **Instrucciones de cómo usar:** **Escaneos de limpieza / OCR** **(Reconocimiento óptico de** **caracteres)**

**¿Para que sirve?:** OCR es una tecnología que reconoce el texto dentro de las imágenes y lo convierte en texto real y seleccionable. Esto hace que los documentos escaneados sean fáciles de buscar, editables y accesibles.

**Paso 1:** Se añade el archivo que quieras comprimir en la casilla, puedes hacer clic en la casilla o arrastrar el archivo a ella.



 Escaneos de limpieza / OCR  
(Reconocimiento óptico de caracteres)

Click o Arrastrar & Soltar

Seleccionar los idiomas que se detectarán en el PDF (Los enumerados son los detectados actualmente):

- Español
- Inglés
- Alemán
- Chino (Simplificado)
- Francés
- Portugués

Modo OCR  
Ignorar las páginas que tienen texto interactivo, solo las páginas OCR que son imágenes

Tipo de procesamiento  
HOOCR (High-Level Optical Character Recognition para alfabeto latino)


Opciones OCR

- Incluir texto en archivo adjunto
- Enderezar archivo de entrada
- Limpiar archivo de entrada
- Limpiar archivo de salida
- Eliminar imágenes del PDF de salida

[Procesar PDF con OCR](#)

**Paso 2:** Eliges el idioma en el que esté el documento que has subido.



 Escaneos de limpieza / OCR  
(Reconocimiento óptico de caracteres)

Click o Arrastrar & Soltar

Seleccionar los idiomas que se detectarán en el PDF (Los enumerados son los detectados actualmente):

- Español
- Inglés
- Alemán
- Chino (Simplificado)
- Francés
- Portugués

Modo OCR  
Ignorar las páginas que tienen texto interactivo, solo las páginas OCR que son imágenes

Tipo de procesamiento  
HOCR (High-Level Optical Character Recognition para alfabeto latino)


Opciones OCR

- Incluir texto en archivo adjunto
- Enderezar archivo de entrada
- Limpiar archivo de entrada
- Limpiar archivo de salida
- Eliminar imágenes del PDF de salida

Procesar PDF con OCR

**Paso 3:** Eliges las opciones de OCR valorando el resultado que quieras obtener.



 Escaneos de limpieza / OCR  
(Reconocimiento óptico de caracteres)

Click o Arrastrar & Soltar

Seleccionar los idiomas que se detectarán en el PDF (Los enumerados son los detectados actualmente):

- Español
- Inglés
- Alemán
- Chino (Simplificado)
- Francés
- Portugués

Modo OCR  
Ignorar las páginas que tienen texto interactivo, solo las páginas OCR que son imágenes

Tipo de procesamiento  
HOOCR (High-Level Optical Character Recognition para alfabeto latino)

Opciones OCR

- Incluir texto en archivo adjunto
- Enderezar archivo de entrada
- Limpiar archivo de entrada
- Limpiar archivo de salida
- Eliminar imágenes del PDF de salida

[Procesar PDF con OCR](#)

## Opciones avanzadas de OCR

En la parte inferior de la herramienta, encontrarás una serie de casillas que permiten mejorar el procesamiento del documento. Dependiendo de la calidad de tu PDF original (especialmente si es un escaneo antiguo), marcar estas opciones optimizará el resultado:

- **Incluir texto en archivo adjunto:** Además de incrustar el texto oculto en el PDF para poder buscar palabras, esta opción recopila todo el texto reconocido y lo adjunta como un archivo de texto plano independiente. Es la opción perfecta si tu objetivo es extraer el contenido masivamente para editarlo en procesadores de texto como LibreOffice Writer.
- **Enderezar archivo de entrada:** Si el documento proviene de una fotografía o un escaneo rápido y las páginas han quedado torcidas, esta función detecta la orientación de las líneas y rota la página automáticamente para dejarla perfectamente horizontal. Esto no solo mejora la presentación, sino que reduce enormemente los errores de lectura del OCR.
- **Limpiar archivo de entrada:** Aplica un filtro digital al documento original *antes* de que el motor empiece a leerlo. Su objetivo es eliminar el "ruido

visual", como fondos grisáceos, manchas de la fotocopidora o sombras del papel. Al darle al sistema una imagen más limpia, la precisión del reconocimiento de texto aumenta.

- **Limpiar archivo de salida:** A diferencia de la anterior, esta opción se aplica al documento final. Mejora el contraste visual del PDF que vas a descargar, intentando blanquear los fondos y oscurecer los caracteres para que la lectura en pantalla o la posterior impresión sea mucho más cómoda y nítida.
- **Eliminar imágenes del PDF de salida:** Si marcas esta casilla, el sistema descartará cualquier fotografía, gráfico, logotipo o ilustración gráfica que contenga el documento original. El PDF resultante contendrá única y exclusivamente el texto reconocido. Es ideal para reducir drásticamente el peso del archivo cuando la información gráfica no es relevante.