

PROGRAMAS DE OCR GRATUITOS

1. Introducción

En este documento se evalúan algunas soluciones de reconocimiento de texto que, cumpliendo con unos requisitos mínimos de calidad y accesibilidad, sean gratuitas. El objetivo final es facilitar a los usuarios información sobre herramientas que, sin coste, les permitan llevar a cabo tareas puntuales de reconocimiento con un nivel mínimo aceptable. En base a ello se han analizado diferentes OCR del mercado, tanto en entorno Web (online) como aplicaciones de escritorio.

Aunque la mayoría de las aplicaciones contemplan el español como uno de los idiomas de reconocimiento, su interfaz no lo soporta (se acompaña imagen en cada aplicación). También se ha comprobado que algunos de los programas a los que se accede en búsquedas realizadas en internet con el parámetro de gratuitos en realidad son versiones con limitaciones funcionales o temporales.

Las condiciones de accesibilidad, funcionalidad y calidad en los resultados obtenidos pueden variar mucho en intervalos de tiempo cortos para las soluciones basadas en servicios web. Se ha apreciado, por ejemplo, que direcciones que antes no eran accesibles por tener limitado su acceso con un captcha, a partir de un determinado momento ya no lo tienen, por lo que las páginas pasan a ser, al menos, usables si no accesibles. También ocurre que herramientas de reconocimiento que brindan buenos resultados con documentos de calidad media en un formato concreto, no son aprovechables para esos mismos documentos en otro formato distinto. Por ejemplo, la misma página evaluada con un mismo OCR online en formato JPG y en formato PDF.

A continuación se muestran los OCRs analizados, indicando algunas de sus características más destacadas, aunque ninguno de ellos cumple bien con la totalidad de requisitos mínimos exigidos. En general el formato y diseño del archivo original se pierden durante el proceso de OCR con la mayoría de aplicaciones analizadas, se obtiene texto plano. Como anexo al documento se ha incluido toda la información resumida en formato tabla.

2. OCRs online (WEB)

Son servicios en línea, por lo que no requieren instalar nada en el ordenador pero sí necesitan disponer de acceso a internet. El proceso en estos casos consiste en acceder a su página web, facilitar en ella el archivo a convertir y a continuación -una vez procesado- se recibe el resultado en texto, en la mayoría de los casos a través de un enlace para descargar, y en algunos también visualizados sobre la misma página web.


Un aspecto muy relevante es la accesibilidad de las páginas web de cada aplicación.

2.1. Free OCR

<http://www.free-ocr.com/es.html>

- OCR gratuito con limitaciones: Tamaño máximo de los ficheros 6 MB.
- Contempla el idioma español para la aplicación.
- Incluye español en los idiomas de conversión, sin embargo en el texto obtenido no aparecen acentos.
- Free OCR acepta archivos JPG, GIF, TIFF BMP y PDF (solo la primera página).
- Accesibilidad: Es posible utilizarlo con revisor de pantalla.
- Calidad reconocimiento: para un mismo contenido en formato JPG y en formato PDF se detecta una respuesta media/aceptable con JPG y muy deficiente con el formato PDF.

Free OCR



The screenshot shows the Free OCR website interface. At the top, there is a navigation bar with links: Inicio, Acerca de, Privacidad, Contacto, Donaciones, and language options (en, de, es, fr). The main content area is titled 'Bienvenido a Free OCR' and describes the service as a free online tool for OCR. It mentions that the service is free and does not require registration. Below this, there is a section for uploading a file, with a text input field and a button labeled 'Examinar...'. To the right of the upload section, there is a list of supported languages for conversion, including Bulgarian, Catalan, Czech, Danish, Dutch, English (checked), Finnish, French (checked), German, Greek, Hungarian, Indonesian, Italian, Latvian, Lithuanian, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian (checked), Spanish (checked), Swedish, Tagalog, Turkish, Ukrainian, and Vietnamese. At the bottom of the language list, there is a button labeled 'Comenzar'. On the right side of the page, there is a sidebar with various links, including 'Traducción Documentos', 'OCR Data Capture Software', 'Curso de Hebreo En Línea', 'Traducción Científica', 'Consilium Benelux DCS bv', 'Curso Bíblico', 'Seguro Adeslas 12 €/mes', 'Cursos Sanidad 2015', 'Comenzar a Descargar', and 'Traducción simultánea'.

2.2. OcrGeek.com

<http://www.ocrgeek.com/es/>

- OCR gratuito.
- Contempla el idioma español para la aplicación, sin embargo se observa que al avanzar en el proceso del OCR y cambiar de página, ésta pasa a estar en inglés
- Incluye español en los idiomas de conversión.
- El resultado del reconocimiento se ofrece solo en formato txt.
- Accesibilidad: Baja. Utiliza una ventana superpuesta que en muchos casos no es detectada por el revisor de pantalla.

OcrGeek.com - Servicio OCR gratuito y online

versión Beta. Powered by DjVu-PDF.com

Condiciones | Condiciones de uso | FAQ | 

Grado en Ingeniería Informática 100% ONLINE  **¡Infórmate!**

Kia Sportage desde 15.200 € **>> Descúbrelo**

Abc

Idioma:
 (+) Añadir mas archivos
 Detectar el idioma...

Formatos: .jpg, .png, .tiff, .bmp, .gif, .pdf, .djvu

Tu dominio .ES a precio imbatible
 ¡OFERTA LIMITADA!
0,29 €/1er año
¡REGISTRA AHORA!
HOSTALIA.com

2.3. OnlineOCR.net

<http://www.onlineocr.net>

- El proceso de reconocimiento incluye un captcha sin alternativa accesible. Por tanto es un OCR no accesible para personas que trabajen con revisor de pantalla.

ONLINE OCR

FREE ONLINE OCR SERVICE

Use Optical Character Recognition software online. Service supports 46 languages including Chinese, Japanese and Korean

CONVERT SCANNED PDF TO WORD

Extract text from PDF and images (JPG, BMP, TIFF, GIF) and convert into editable Word, Excel and Text output formats

1 STEP - Upload file **2 STEP - Select language and output format** **3 STEP - Convert**

Select file...

HCR_Aragón.pdf

Incorrect Captcha code

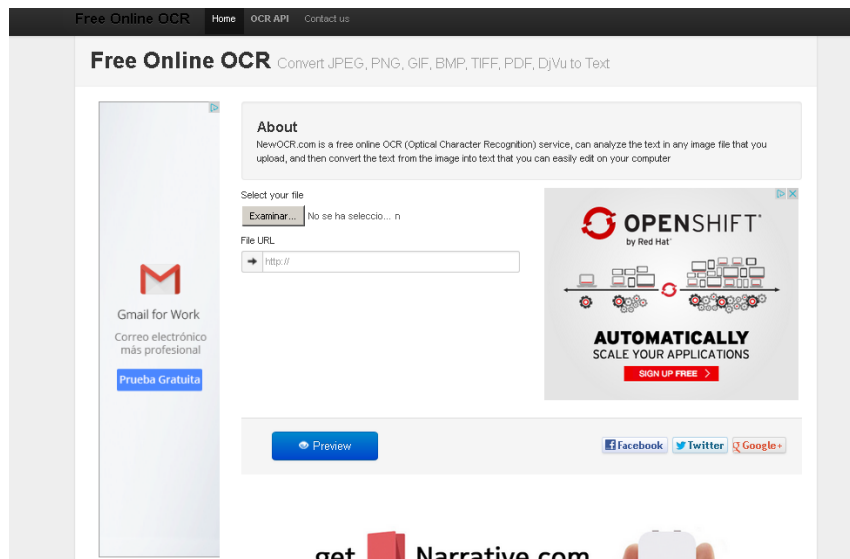
Use OCR software

2.4. NewOCR.com

<http://www.newocr.com/>

- OCR gratuito.
- No contempla el idioma español para la interfaz de la aplicación.

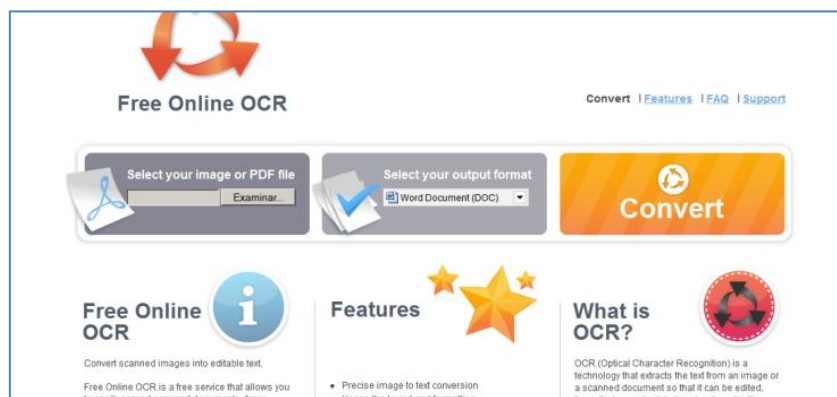
- Permite exportar a múltiples formatos (doc, txt, etc.), y presenta el resultado en la propia página web antes de descargarlo.
- Basado en el OCR Tesseract. Los resultados del reconocimiento han sido mejores que con la mayoría de herramientas online analizadas.
- Accesibilidad/usabilidad. Las páginas web de esta aplicación son difíciles de manejar con un revisor de pantalla, resultando tedioso su uso.



2.5. Free Online OCR

<http://www.free-online-ocr.com/>

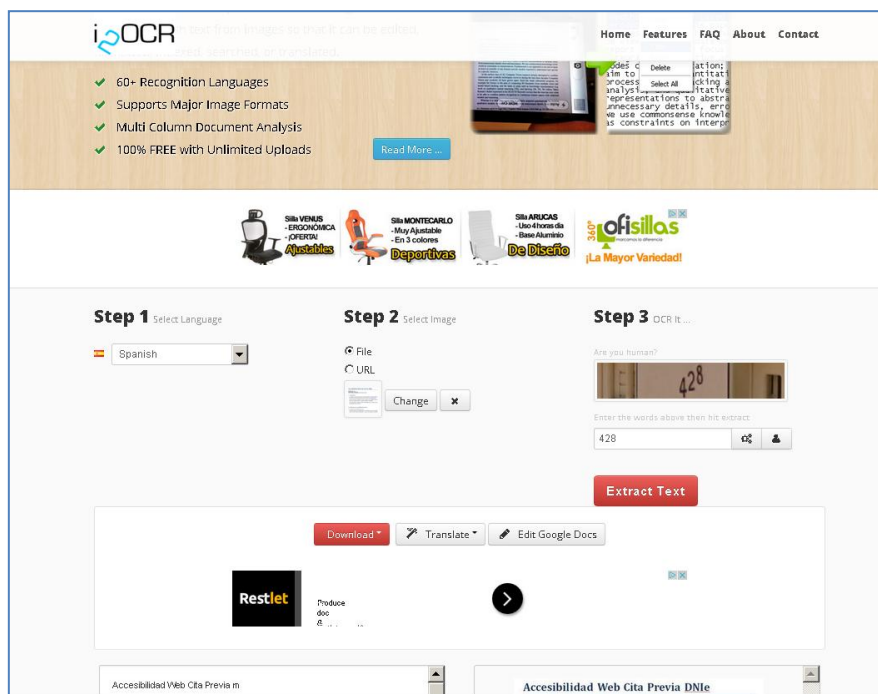
- OCR gratuito.
- No incluye captcha, ni requiere registrarse.
- Interfaz en inglés.
- No reconoce acentos al convertir documentos en español.
- Páginas con muchas dificultades de accesibilidad, en especial es complicado localizar el botón que ejecuta la conversión con un revisor de pantalla.



2.6. I2OCR

www.i2ocr.com

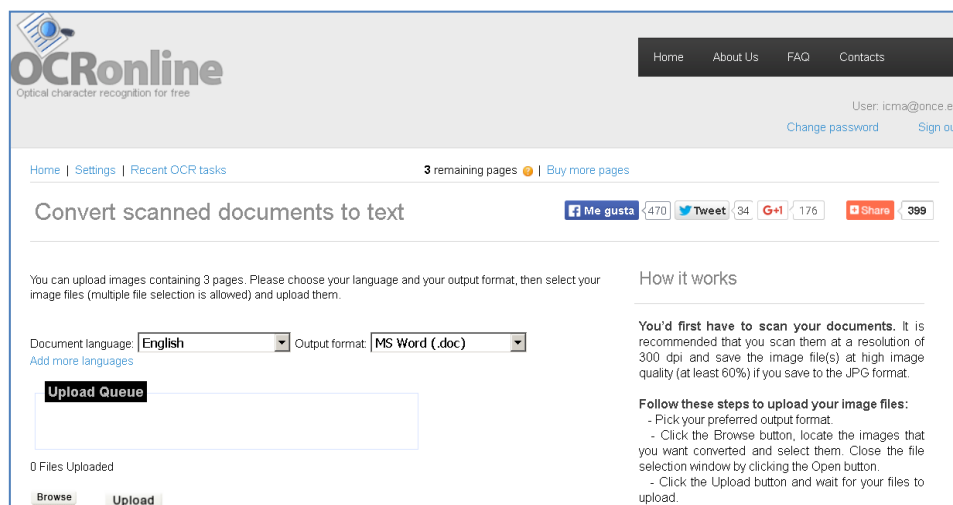
- Gratuito
- Interfaz en inglés
- Captcha visual y sonoro (solo en inglés)
- Páginas muy difíciles para seguir con revisor de pantalla. Por ejemplo, se detecta que al seleccionar el nombre del documento imagen la página web se recarga automáticamente y desplaza el revisor de pantalla al inicio de página constantemente.



2.7. OCR Online

<http://www.ocronline.com/>

- Gratuito parcialmente: 5 páginas por semana.
- Requiere registrarse previamente.
- Sus páginas web presentan problemas importantes de accesibilidad que impiden su uso con revisores de pantalla. (Por ejemplo, no están etiquetados los botones de acceso a ejecutar la conversión del archivo).



2.8. ABBYY FineReaderOnline

<http://finereaderonline.com/en-us>

- Admite 10 páginas por mes gratuitas.
- Incluye un captcha sin alternativa accesible, por lo que este OCR debe considerarse no accesible para usuarios de revisor de pantalla.



3. Aplicaciones de OCR de escritorio

Estas soluciones implican descargar e instalar una aplicación en el ordenador. Al instalar este tipo de software se debe estar atento ya que se incluyen instalaciones de programas de terceros que en realidad no es necesario descargar ni instalar.

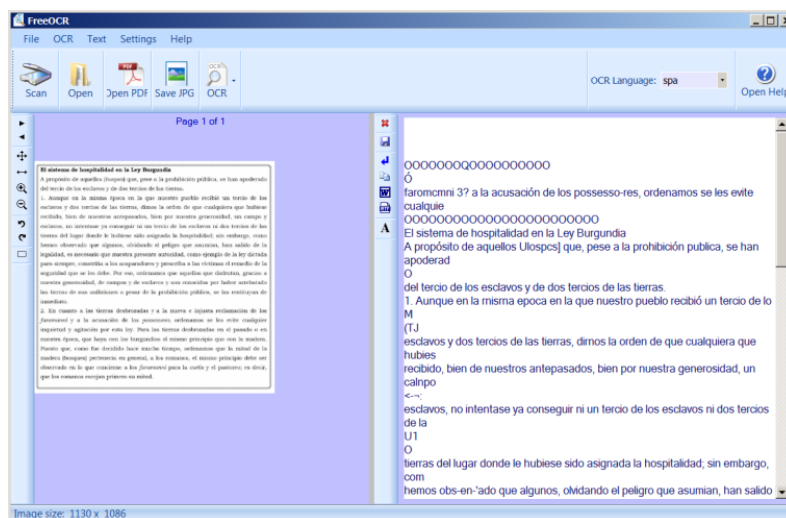
3.1. Simpleocr 3.1

- Esta aplicación no admite el idioma español ni en la interfaz de la aplicación ni como idioma de reconocimiento.
- No soporta archivos PDF.



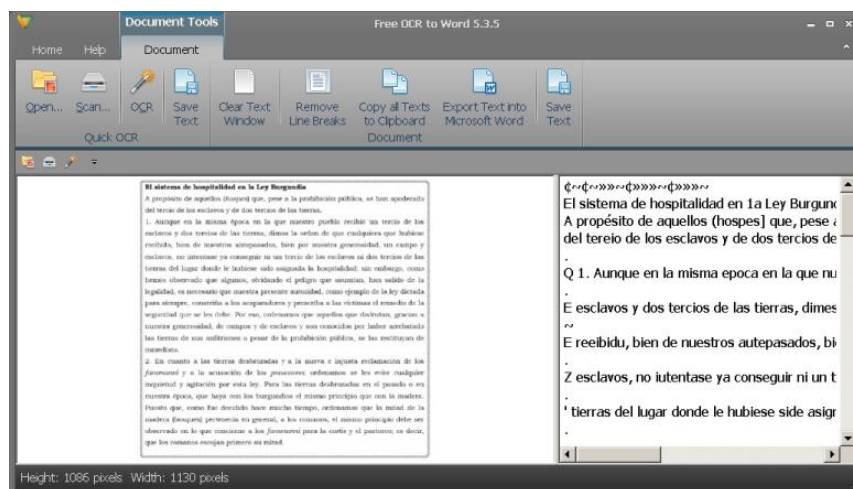
3.2. FreeOcr.net -Versión 5.41 (marzo 2015)

- Aplicación gratuita.
- Permite importar contenidos desde escáneres TWAIN, ficheros PDF y formatos de imagen de uso habitual.
- Este OCR está basado en Tesseract v3.
- La interfaz de la aplicación no está disponible en español.
- Contempla español en los idiomas de reconocimiento.
- Aplicación accesible con revisor de pantalla, aunque su manejo es complejo.
- Calidad del reconocimiento: regular con ficheros JPG y baja con ficheros PDF.
- La aplicación puede descargarse desde: <http://www.paperfile.net/>.
- Este OCR también puede localizarse en internet con el nombre de *SoftfreeOcr*.



3.3. Free OCR to Word 7.6.5

- Aplicación gratuita con interfaz en inglés.
- Contempla español para reconocimiento del texto si bien se observa que el texto convertido carece de acentos.
- Es compatible con diversos tipos de archivos gráficos, y escáneres. No admite archivos PDF.
- Es accesible con JAWS solo utilizando teclas rápidas. La barra de menú en formato cinta no es accesible.



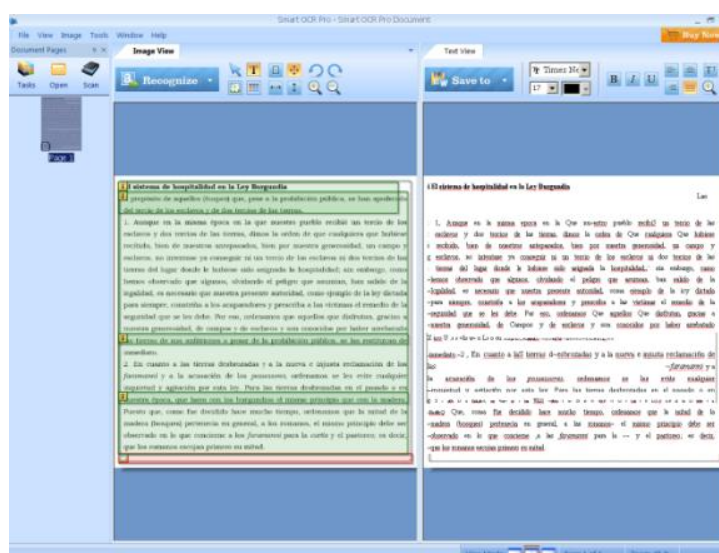
3.4. Cuneiform OCR

- Aplicación gratuita.
- La interfaz está en inglés y contempla español para reconocimiento del texto.
- Se puede trabajar con revisor de pantalla.
- Calidad de reconocimiento: baja.



3.5. Smart OCR Pro

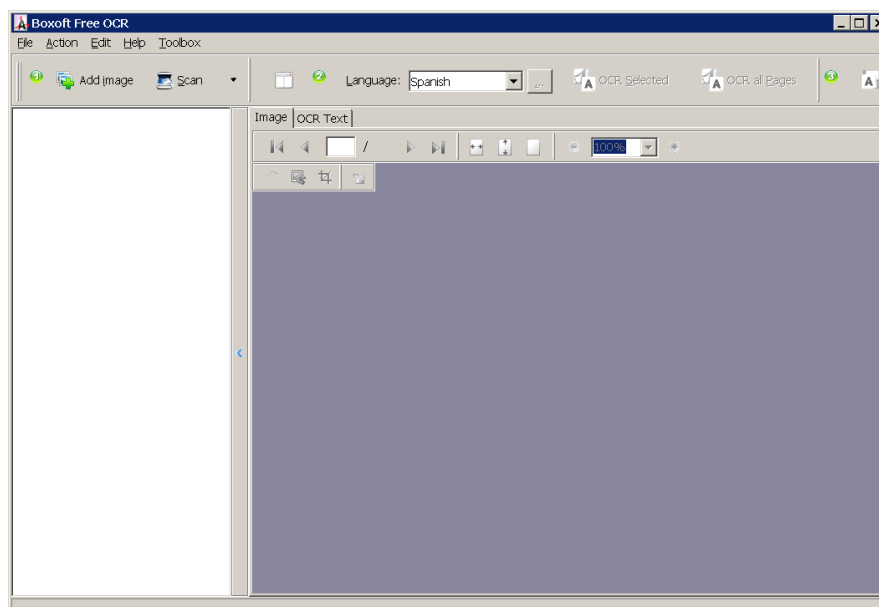
- La interfaz está en inglés, contempla español para reconocimiento del texto.
- Además de la aplicación de escritorio, se añade una nueva pestaña en Microsoft Word cuyo manejo con JAWS es más simple.
- No es gratuita; se puede utilizar en modo demo. La limitación de la demo es en tiempo.



3.6. Boxoft Free OCR (freeware) 1.1

- Aplicación no totalmente gratuita. Aparecen atenuadas funciones relacionadas con el proceso de OCR.
- Interfaz en inglés.

- Admite español como idioma de conversión.
- Dirección descarga:
<http://www.softpicks.es/software/Negocios/Herramientas-y-Programas-para-Oficina/Boxoft-Free-OCR-freeware-es-197714.htm>



4. Conclusiones

En función de los datos anteriores se puede concluir que, en general, las aplicaciones que son gratuitas, accesibles y contemplan el idioma español para el reconocimiento no disponen de interfaz en español y, además, la calidad del reconocimiento es baja.

No se ha encontrado ninguna aplicación que sea claramente recomendable para la finalidad perseguida, entendiendo que los parámetros de mayor peso para la selección sean la gratuidad y la accesibilidad.

A nivel particular, pudiera ser que alguna de las aplicaciones en las que se menciona que sí permiten trabajar con un revisor de pantalla -aunque no de forma simple y sencilla- sea aplicable para alguna situación puntual pero, desde luego, no como recomendación de propósito general.

ANEXO 1 - TABLA DE OCR ONLINE (WEB)

OCR	URL	PRECIO	IDIOMA INTERFAZ	CAPTCHA	OBSERVACIONES
FreeOCR	http://www.free-ocr.com/es.html	Gratuito Limitaciones: Tamaño máx. ficheros 6Mb.	Admite español e inglés por defecto.	Reciente- mente han suprimido el captcha de su página.	- Es posible utilizarlo con revisor de pantalla. - Aunque contempla el idioma español en el texto obtenido no aparecen acentos. - Calidad de reconocimiento: muy baja con formatos PDF y baja con formatos JPG.
OcrGeek.com	http://www.ocrgEEK.com/es/	Gratuito	Puede ponerse en ESPAÑOL, pero al cambiar de página la interfaz vuelve a inglés.		- El resultado del reconocimiento se ofrece solo en formato txt, PDF o DjVu. - Accesibilidad: Baja. Utiliza una ventana superpuesta que en muchos casos no es detectada por el revisor de pantalla
OnlineOCR.net	http://www.online-ocr.net/	Gratuito Tamaño máx. archivos 5Mb	No admite español	Tiene un captcha no accesible	Aplicación NO ACCESIBLE
NewOCR.com	https://www.newocr.com/	Gratuito	No admite español	—	- Accesibilidad/ usabilidad: páginas web difíciles de manejar con un revisor de pantalla, resultando tedioso su uso. - Los resultados del reconocimiento han sido mejores que con la mayoría de herramientas online analizadas

Free OnlineOCR	http://www.free-online-ocr.com/	Gratuito	No admite español	—	- Accesibilidad: baja - No reconoce acentos. - Mantiene el formato y disposición original del texto. - El proceso de conversión es lento y la calidad de reconocimiento es media.
i2OCR	http://www.i2ocr.com/	Gratuito	No admite español	Tiene captcha gráfico con alternativa sonora (solo en inglés)	-Accesibilidad: mala. Captcha sonoro solo en inglés, páginas que se recargan automáticamente, etc.
OCR Online	http://ocronline.com/	Gratuito parcialmente: 5 páginas por semana	No admite español	—	Aplicación NO ACCESIBLE con revisor de pantalla. Requiere registrarse previamente.
ABBY Fine Reader Online	http://www.ocronline.com/	Gratuito parcialmente: 10 páginas por mes	No admite español	Tiene un captcha no accesible	Aplicación NO ACCESIBLE

ANEXO 2 - TABLA APLICACIONES DE OCR DE ESCRITORIO GRATUITAS

NOMBRE	PRECIO	IDIOMA INTERFAZ	IDIOMA CONVERSIÓN	OBSERVACIONES
SimpleOCR	Gratuito	No admite español	No contempla español	<ul style="list-style-type: none"> - OCR no útil para textos en español. - No soporta archivos PDF.
FreeOcr.net - Versión 5.41 (Marzo 2015).	Gratuito	No admite español	Contempla español	<ul style="list-style-type: none"> - Permite importar contenidos desde escáneres TWAIN, ficheros PDF y formatos de imagen de uso habitual. - Puede utilizarse con revisor de pantalla pero su manejo no es simple. - Calidad del reconocimiento: regular con ficheros JPG y baja con ficheros PDF.
Free OCR to Word 7.6.5	Gratuito	No admite español	Contempla español si bien se observa que el texto convertido carece de acentos	<ul style="list-style-type: none"> - Es compatible con diversos tipos de archivos gráficos y escáneres. No admite archivos PDF. - Es accesible con JAWS solo utilizando teclas rápidas. La barra de menú en formato cinta no es accesible.
Cuneiform OCR	Gratuita	No admite español		<ul style="list-style-type: none"> - No es compatible con archivos PDF. - Se puede utilizar con revisor de pantalla. - Calidad reconocimiento: baja.
Boxoft Free OCR 1.1	No gratuita	No admite español	Contempla español	<ul style="list-style-type: none"> - Una vez instalada aparecen atenuadas opciones del proceso de OCR que impiden utilizar esta aplicación.

Madrid, 4 de enero de 2016

ONCE-CIDAT