



ARCHIVO
GENERAL
DE LA NACIÓN
COLOMBIA



MINCULTURA



TODOS POR UN
NUEVO PAÍS
PAZ EQUIDAD EDUCACIÓN

Manual de Instalación

SOFTWARE
gimageReader
3.2.99



Grupo de Innovación y apropiación de
Tecnologías de la Información Archivística

Compilador: Pedro Antonio Gómez Guarín

2018



Introducción

gImageReader es una herramienta diseñada para el reconocimiento y la extracción de los caracteres contenidos en un documento ya sea en formato PDF o una imagen, mediante la técnica OCR (siglas en inglés para Reconocimiento Óptico de Caracteres).

gImageReader es una interfaz gráfica para la herramienta de software libre Tesseract-OCR, que es un motor de Reconocimiento Óptico de Caracteres (OCR) Inicialmente desarrollado por HP Labs como software propietario, fue publicado como código abierto en el año 2005 y desde el año 2006, Google se ha hecho cargo de su desarrollo y está disponible para múltiples sistemas operativos.

Página del proyecto: <https://opensource.google.com/projects/tesseract>

Como principales características este programa puede:

- Importar documentos e imágenes PDF desde el disco, dispositivos de escaneo, portapapeles y capturas de pantalla
- Procesar imágenes y documentos múltiples de una sola acción
- Definir el área de reconocimiento manual o automático en el documento
- Reconocer a texto plano o a documentos hOCR
- Mostrar el texto reconocido directamente al lado de la imagen que se está procesando
- Postprocesamiento del texto reconocido incluyendo la revisión ortográfica
- Genera documentos PDF a partir de documentos hOCR
- Licencia código abierto y libre distribución.

Para realizar este manual utilizamos la versión de gImageReader 3.2.99 para Windows

Requisitos

Idiomas: Aunque el procedimiento de instalación, la página Web e incluso el programa se encuentran en inglés, si al realizar la instalación de Tesseract-OCR instalamos el paquete de idioma español, el programa será capaz de reconocer correctamente textos en este idioma.

Instalar primero Tesseract-OCR

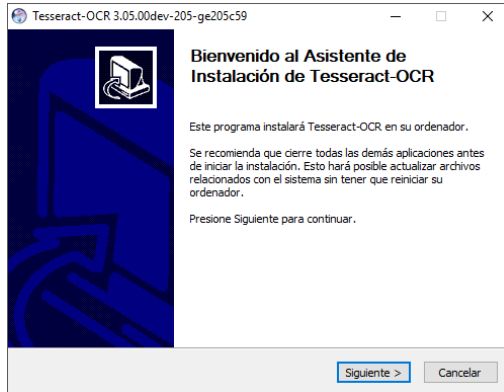
Antes de instalar gImageReader se debe instalar el motor de reconocimiento OCR, para esto debe acceder a la Web oficial para descargar los componentes de [Tesseract-OCR](#) 3.05.02 que funciona con sistemas operativos a 64 bits

Descarga directa:

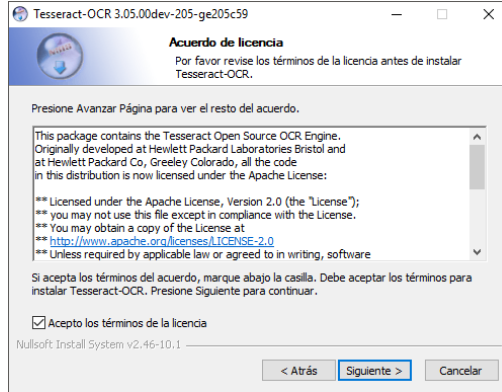
https://es.osdn.net/projects/sfnet_tesseract-ocr-alt/downloads/tesseract-ocr-setup-3.02.02.exe/



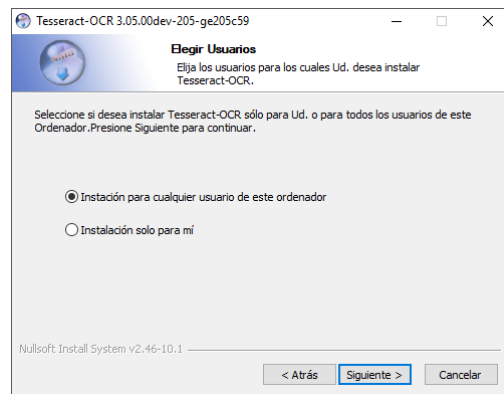
Inicio de instalación



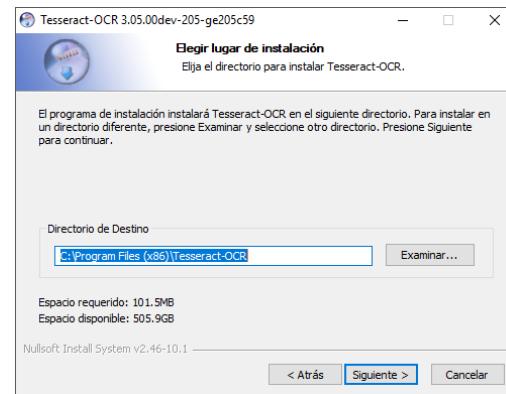
Aceptar términos



Tipo de instalación usuarios



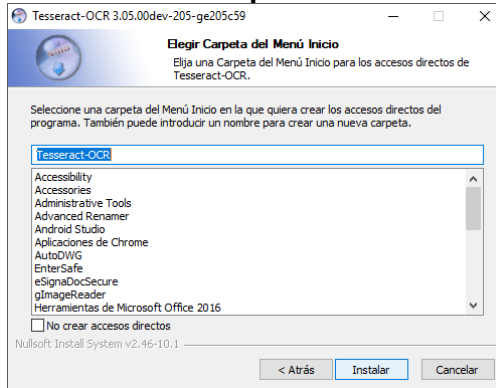
Lugar de instalación



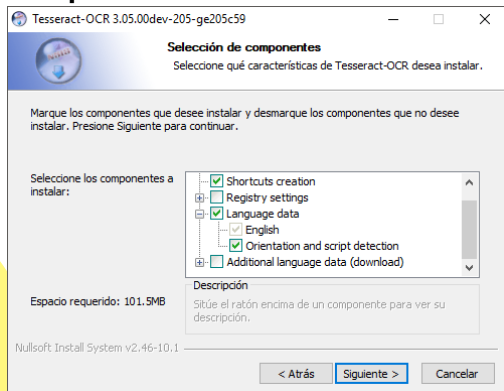
Idioma recomendado español



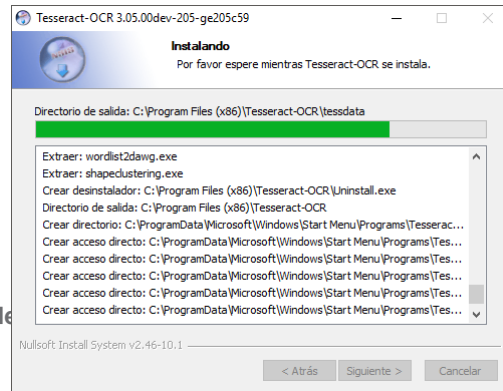
Nombre de la carpeta



Seleccionar los componentes a instalar



Finalizando la instalación

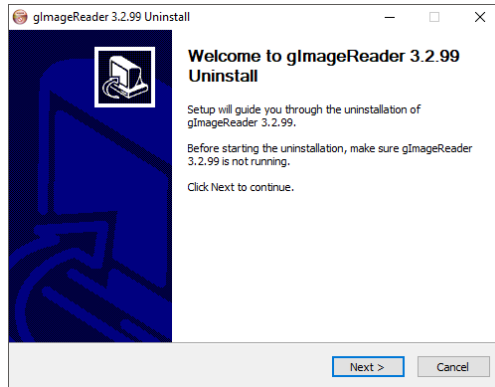


Instalar gImageReader 3.2.99

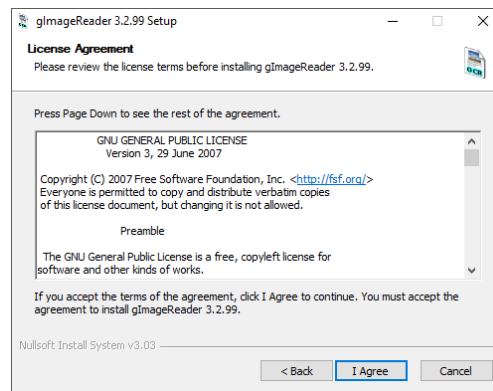
Después de instalar **Tesseract-OCR** procedemos a la instalación de **gImageReader 3.2.99**, para ello necesitamos el programa instalador que lo podemos descargar del siguiente link:

<https://sourceforge.net/projects/gimagereader/files/>

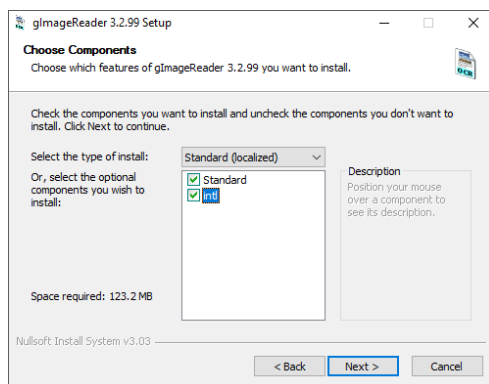
Inicio de instalación



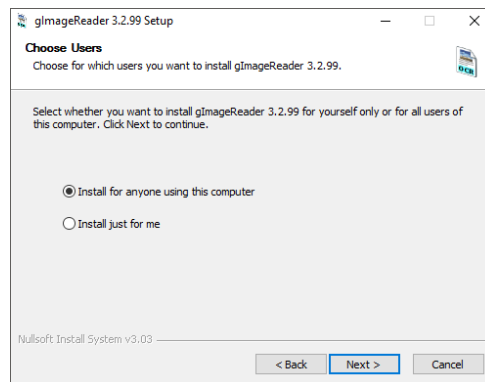
Licencia



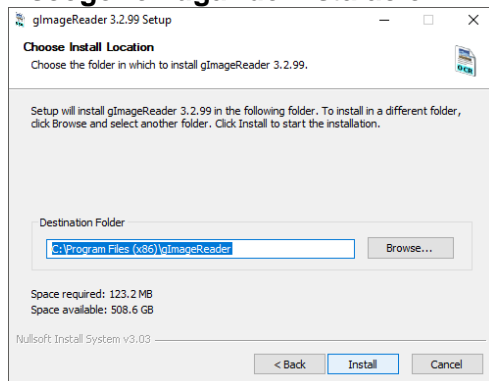
Componentes a instalar



Tipo de instalación usuarios



Escoger el lugar de instalación



Instalación completa

