

<http://www.paperfile.net/>:

FreeOCR (paperfile.net)

Contents

About FreeOCR.....	2
Scanning Software.....	2
OCR Engine.....	2
License.....	2
Download FreeOCR.....	3
FreeOCR v5.4.1.....	3
Download FreeOCR.....	3
Enhancements to latest version (5.4.1).....	3
All Enhancements to Version 5 (5.1.2 - 5.3.3).....	4
All Enhancements to Version 4 (4.0.1 - 4.7.0).....	4
Uninstall Info.....	4
Downloading FreeOCR.....	5
AbleWord the Free PDF Editor.....	5
How to Use and Install FreeOCR Languages.....	6
FreeOCR includes the following languages by default.....	6
How to download and install additional languages.....	6
Getting Started with FreeOCR.....	8
First open a document.....	8
OCR.....	8
OCR Selection.....	9
How to OCR a document containing images and columns.....	9



About FreeOCR

FreeOCR is a free Optical Character Recognition Software for Windows and supports scanning from most Twain scanners and can also open most scanned PDF's and multi page Tiff images as well as popular image file formats. FreeOCR outputs plain text and can export directly to Microsoft Word format.

Free OCR uses the latest Tesseract (v3.01) OCR engine. It includes a Windows installer and It is very simple to use and supports opening multi-page tiff documents, Adobe PDF and fax documents as well as most image types including compressed Tiff's which the Tesseract engine on its own cannot read .It now can scan using Twain and WIA scanning drivers.

FreeOCR V4 includes Tesseract V3 which increases accuracy and has page layout analysis so more accurate results can be achieved without using the zone selection tool.

Scanning Software

As well as OCR FreeOCR can scan and save images as JPG's and we are currently working on "Scan to PDF" capability with the option to save as searchable PDF

OCR Engine

The included Tesseract OCR PDF engine is an open source product released by Google. It was developed at Hewlett Packard Laboratories between 1985 and 1995. In 1995 it was one of the top 3 performers at the OCR accuracy contest organized by University of Nevada in Las Vegas. The Tesseract engine source code is now maintained by Google and the project can be found here: <http://code.google.com/p/tesseract-ocr/>

License

FreeOCR is a freeware OCR & scanning software and you can do what you like with it including commercial use. The included Tesseract OCR engine is distributed under the Apache V2.0 license.

[Terms and Conditions](#) - [Privacy Policy](#)

Download FreeOCR

FreeOCR is Optical Character Recognition Software for Windows and supports scanning from most Twain scanners and can also open most scanned PDF's and multi page Tiff images as well as popular image file formats.

FreeOCR outputs plain text and can export directly to Microsoft Word format.

FreeOCR v5.4.1

System Requirements

Operating System

Windows XP (Service pack 2)

Windows Vista

Windows 7

Windows 8 Desktop (not RT/Metro)

Recommended Minimum Specification

1024 MB Memory (RAM)

20 MB Free Disk Space

SVGA Resolution Display

.Net Framework 2.0 or higher



To enable scanning of images you will need a desktop document scanner that uses Twain or WIA compatible scanning drivers.

FreeOCR requires the .Net Framework V2.0 to be installed for XP users. You can download from Softpedia [here](#)

This framework is included with Windows Vista, 7, 8 so only may need installing on XP.

Download FreeOCR

[Download Here](#)

After downloading simply double click and follow the instructions to install.

Enhancements to latest version (5.4.1)

- Testing with Windows 10 (Technical Preview)
- Scanning fixes to makes/models of scanner
- Better PDF compatibility for PDF Open function
- Minor fixes

All Enhancements to Version 5 (5.1.2 - 5.3.3)

- Automatic scan cleanup to improve OCR quality
- Twain scanning fixes to stop freezing on some makes/models of scanner
- Windows 8.1 testing
- Changed automatic updater to direct the user to our download page because the updater.exe was getting flagged as a false positive with a couple of AV engines
- Minor bug fixes

All Enhancements to Version 4 (4.0.1 - 4.7.0)

- Multi page Twain Scanning
- OCR whole document in one go
- Uses Tesseract V3 for higher accuracy and ability to recognize text columns
- Windows 8 compatible
- Automatic updater
- Text post processing
- FreeOCR V4 Now includes the following languages:

Eng - English, Dan - Danish, Deu - German, Fin - Finnish, Fra - French, Ita - Italian, Nld - Dutch, Nor - Norway, Pol - Poland, Spa - Spanish and Swe - Swedish.

Uninstall Info

FreeOCR includes an uninstaller if you wish to remove the software at a later date. This can be accessed from Control Panel - "Add Remove Software" on XP or "Programs and Features" on later operating systems.

[Terms and Conditions](#) - [Privacy Policy](#)

Downloading FreeOCR

Your download should start automatically, if you have any problems then [Click Here](#) to retry.

If you like FreeOCR you may also like:

AbleWord the Free PDF Editor



AbleWord is a very capable PDF Editor and word processing application that can read and write most popular document formats including PDF's. It is fully featured, supporting image formatting, tables, headers & footers and includes spell checking and print preview functions. Best of all it is totally free even for commercial use.

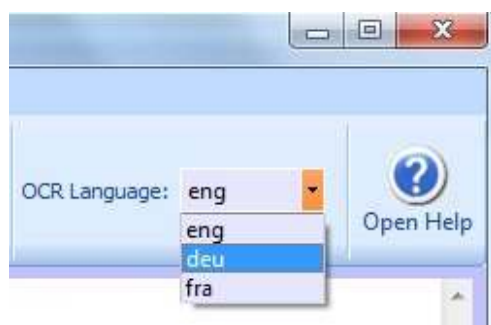
[**Visit AbleWord Site**](#)

How to Use and Install FreeOCR Languages

Free OCR uses the latest Google Tesseract OCR engine so you can install any language that this engine supports.

FreeOCR includes the following languages by default

Eng - English
 Dan - Danish
 Deu - German
 Fin - Finnish
 Fra - French
 Ita - Italian
 Nld - Dutch
 Nor - Norway
 Pol - Polish
 Spa - Spanish
 Swe - Swedish



to use simply select the 3 digit language code in the toolbar before OCR'ing. If you need additional languages then follow the instructions below.

How to download and install additional languages

Visit the [Tesseract download page](#) and download your chosen language pack



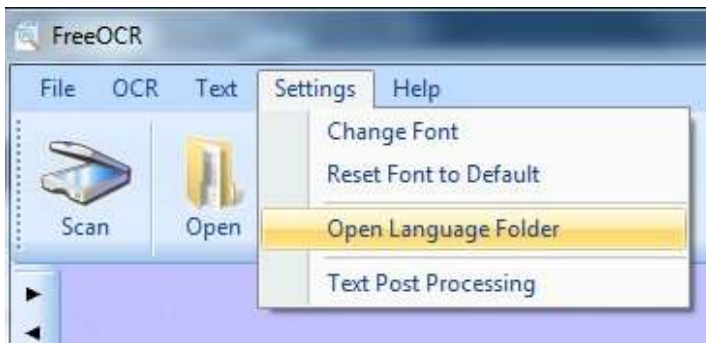
Make sure the language file is for Tesseract 3.00 or higher (the 2.00 files will not work)

After downloading you will need to uncompress the file, we use [7 Zip](#) but [WinRAR](#) or similar programs will work.

You should end up with a file starting with the 3 digit country and ending with **.traineddata**

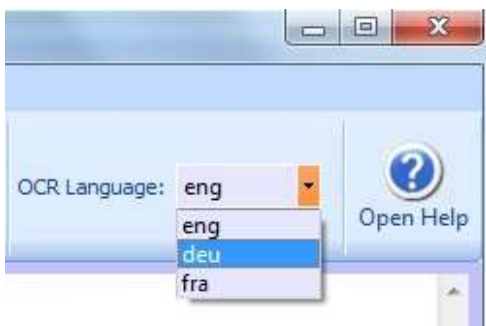
Tesseract uses the ISO 3 letter country codes, [more info here](#)

Now open the data folder for Tesseract.



The data folder will open in Windows explorer. Now just Drag & Drop the language data file into the tessdata folder.

Now if you close and reopen FreeOCR it will see the new language file and you can choose it before starting OCR



If your OCR language does not exist and you would like to train the engine for a new language then the Tesseract project can be found here: <http://code.google.com/p/tesseract-ocr/>

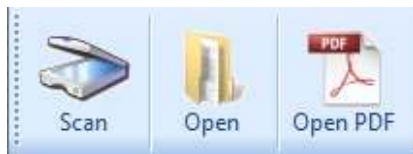
[Terms and Conditions](#) - [Privacy Policy](#)

Getting Started with FreeOCR

FreeOCR supports scanning from most Twain scanners and can also open most scanned PDF's and multi page Tiff images as well as popular image file formats. FreeOCR outputs plain text and can export directly to Microsoft Word format.

FreeOCR is very simple to use and all of the buttons have a description if you hover your mouse over them.

First open a document



Scan - Use to Scan a paper image, this requires a Twain compatible scanner.

Open - Use to load an image such as Tiff, Jpeg, Bmp. FreeOCR supports multi page Tiff documents.

Open PDF - Use to import a scanned PDF document.

OCR

Choose the OCR language you want to use

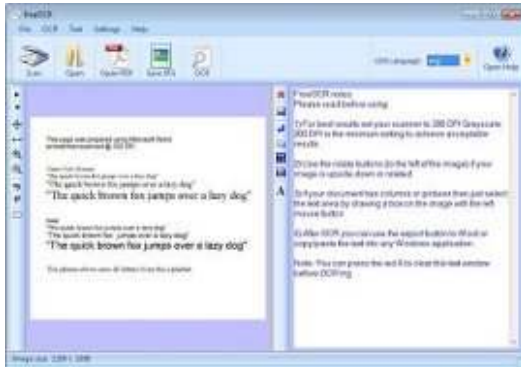


Installed languages include:

- Eng - English
- Dan - Danish
- Deu - German
- Fin - Finnish
- Fra - French
- Ita - Italian
- Nld - Dutch

Nor - Norway
Pol - Polish
Spa - Spanish
Swe - Swedish

Press the OCR button and choose if you want to process the current page or the whole document.



The image is shown on the left and the OCR text on the right

You can copy and paste the text into any other Windows application or press one of the export buttons on the text toolbar.

OCR Selection

You can draw a box around part of the image using your mouse and then OCR the current page. This is handy if you just want to get the text from one region of the page.

How to OCR a document containing images and columns

This is a PDF with a fake UFO report you can download the file [HERE](#) if you want to try (you may need to right-click and choose 'save link as' or 'save target as')



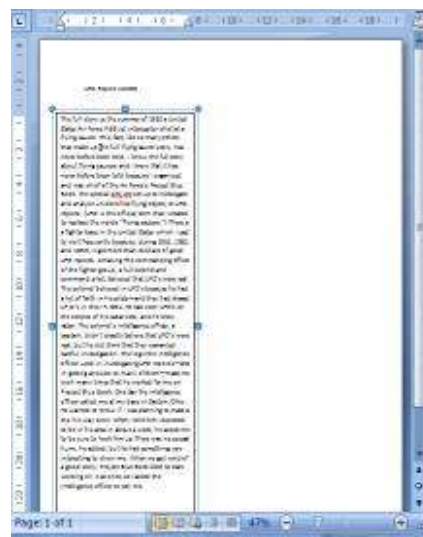
I am going to use MS Word to recreate the document but you could use Open Office or publisher, in fact any program that supports columns and photos.

1) now to start off I am just going to OCR the page title



2) Then press the Export to Word button which will open up a new document with the title in it, if you are not using Word then just open a new document and copy and paste the text.

3) Next I have selected and OCR 'ed the first column also I pressed the 'remove line breaks button' then in Word created a new Text Box (Insert menu on Word) and copy and pasted the text into it as shown below.



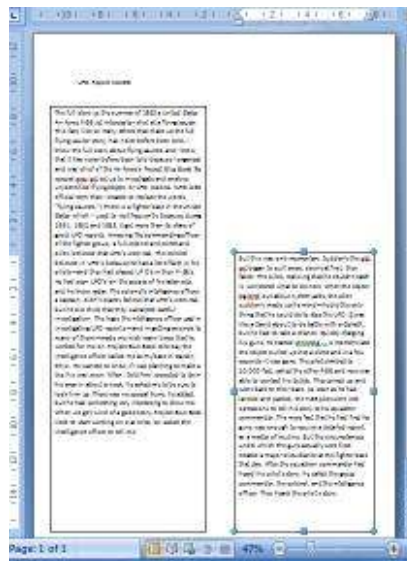
MS Word

FreeOCR

4) Now clear the text window and do the same for the second column

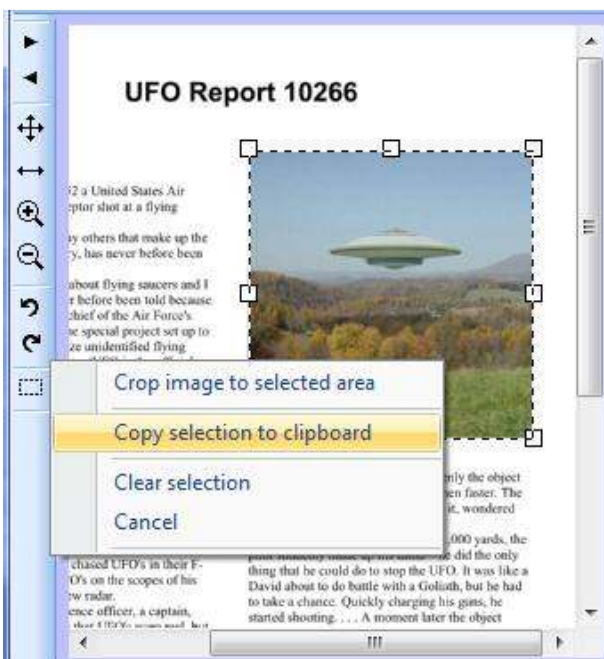


FreeOCR

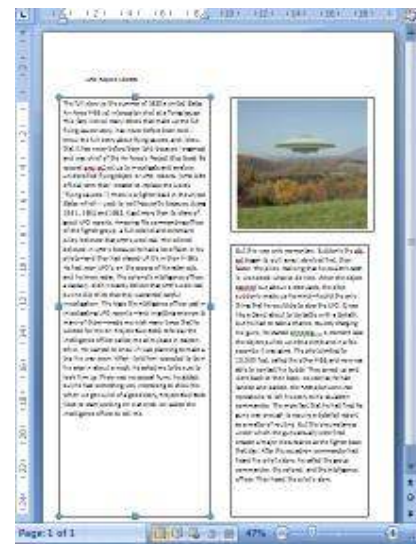


MS Word

5) Now draw a box around the photo to select the image, then press the selection button and choose 'Copy selection to clipboard' this will place the selected image onto the Windows clipboard so now we can just paste the image into Word (Tip: paste the image into a text box then you can position it easily)



FreeOCR



MS Word

Now we have all of the text and images in Word you can play around positioning the text boxes and adjusting the font sizes to suit and don't forget to do a final proof read and spell check.

As you can see there is a manual element of work involved to get the best out of FreeOCR but I am sure you will find it easier than recreating a document from scratch.

I hope you enjoy using FreeOCR

[Terms and Conditions](#) - [Privacy Policy](#)