

HOW TO INSTALL OCR TESSERACT 2.00

Linux Platform

- Download the Tesseract-ocr package at <http://tesseract-ocr.googlecode.com/files/tesseract-2.00.tar.gz>
- Download the Tesseract-ocr language data package you want here: <http://code.google.com/p/tesseract-ocr/downloads/list>
- Choose the different packages you want, named "tesseract-2.00.XX.tar.gz" where "XX" represents the language you want.
- Now open a terminal and go to the folder containing your latest downloads with the command "cd".
- You need to run the next commands to install OCR Tesseract : (\$> represents the prompt of your terminal)

```
$>tar -xvzf tesseract-2.00.tar.gz
```

```
$>cd tesseract-2.00
```

```
$>./configure
```

```
$>make
```

```
$>sudo make install
```

```
$>tar -xvzf tesseract-2.00.XX.tar.gz    ("XX" represents the prefix language)| do it
```

```
$>sudo cp ./tessdata/* /usr/local/share/tessdata/ | for each
```

```
$>rm -rf tessdata | language archive
```

- Now that you have OCR tesseract on your computer, you can try it with these commands:
\$>tesseract phototest.tif test
\$>cat test.txt

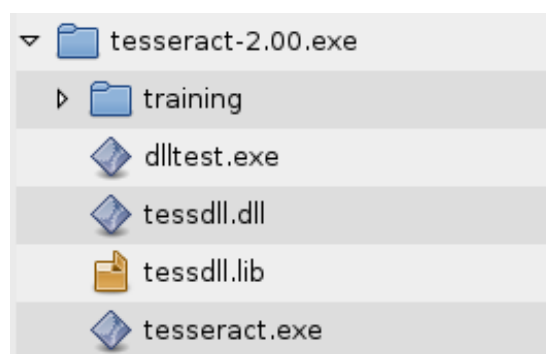
"test.txt" contains the same text than the picture "phototest.tif".

Windows Platform

- Download the Tesseract-ocr package at <http://tesseract-ocr.googlecode.com/files/tesseract-2.00.exe.tar.gz>
- Download the Tesseract-ocr language data package you want here: <http://code.google.com/p/tesseract-ocr/downloads/list>
- Choose the different packages you want, named "tesseract-2.00.XX.tar.gz" where "XX" represents the language you want.
- Create a new folder named "tesseract" where you want, for example at the root of [C:\](#) .
- Now, open the directory where are the latest downloads.
- Extract the content of the first download named "tesseract-2.00.exe.tar.gz" in your new folder "tesseract".

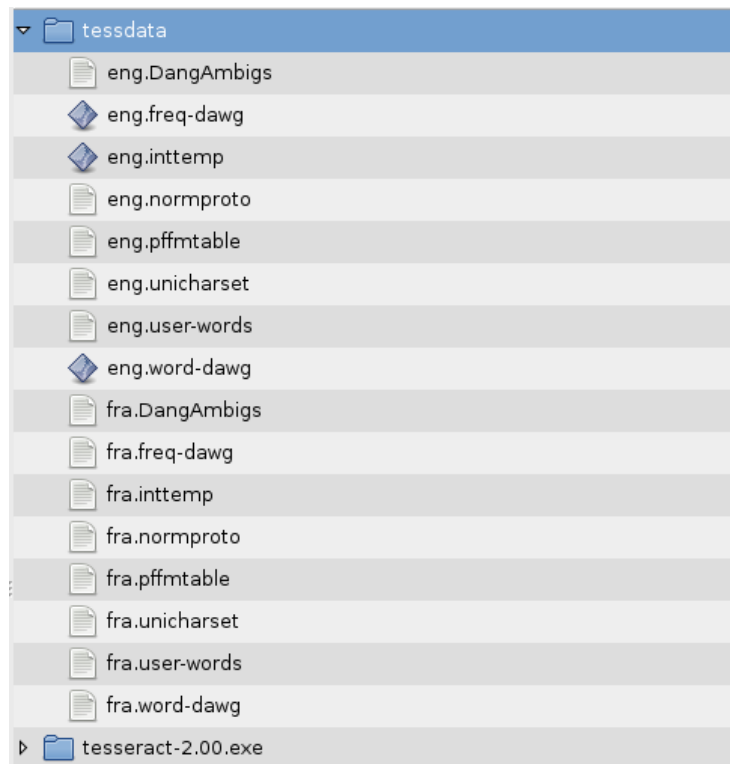
Warning : Be aware that the extract folder name has the ".exe" suffix.

Now you probably have something like this in the folder "tesseract" :



- Create a new folder named "tessdata" in the folder "tesseract" and put in the content of the folder named "tessdata" of each language archive.

Now you probably have something like this in the folder "tesseract" if you got the English and French language archive :



- Now you need to reference the "tesseract.exe" in your path. Modify the Environment variable named "Path", and add "your-path-to-the-tesseract-folder\tesseract-2.00.exe" to the current value.

To do it use the Command line (start → run → type " cmd " and click OK)

(\$> represents the prompt of your terminal)

```
$> set path=%path%;your-path-to-the-tesseract-folder\tesseract-2.00.exe
```

- Create a new environment variable named "TESSDATA_PREFIX" and put for its value "your-path-to-the-tesseract-folder\".

(Right click on "My Computer" → Properties → Advanced → Environment Variables)

- Reboot your computer