

Thank you very much for having an interesting for our product.

【Preparation】

- ① Please download the OCR tool installer provided by Google from the following URL.

<https://github.com/UB-Mannheim/tesseract/wiki>

Home

Stefan Weil edited this page on 28 Nov 2020 - 68 revisions

Tesseract at UB Mannheim

The Mannheim University Library (UB Mannheim) uses Tesseract to perform OCR (optical character recognition) of historical German newspapers (*Allgemeine Preußische Staatszeitung*, *Deutscher Reichsanzeiger*). The latest results with OCR from more than 360,000 scans are available [online](#).

Normally we run Tesseract on Debian GNU Linux, but there was also the need for a Windows version. That's why we have built a Tesseract installer for Windows.

WARNING: Tesseract should be either installed in the directory which is suggested during the installation or in a new directory. The uninstaller removes the whole installation directory. If you installed Tesseract in an existing directory, that directory will be removed with all its subdirectories and files.

The latest installers can be downloaded here:

- [tesseract-ocr-w32-setup-v5.0.0-alpha.20201127.exe](#) (32 bit) and
- [tesseract-ocr-w64-setup-v5.0.0-alpha.20201127.exe](#) (64 bit) resp.

We don't provide an installer for Tesseract 4.1.0 because we think that the latest version 5.0-alpha is better for most Windows users in many aspects (functionality, speed, stability). Version 4.1 is only needed for people who develop software based on the Tesseract API and who need 100 % API compatibility with version 4.0.

ここからどちらかをダウンロードしてください。



Pages 6

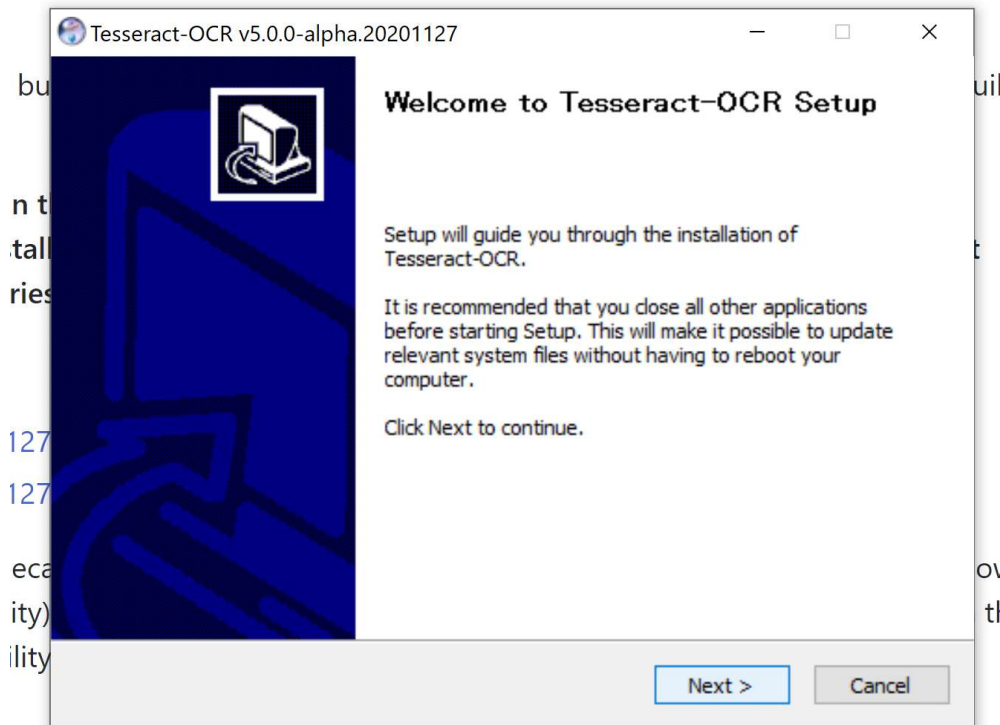
Find a Page...

- Home
- FAQ
- Fonts for Tesseract training
- Install additional language and script models
- Project ideas
- Windows build

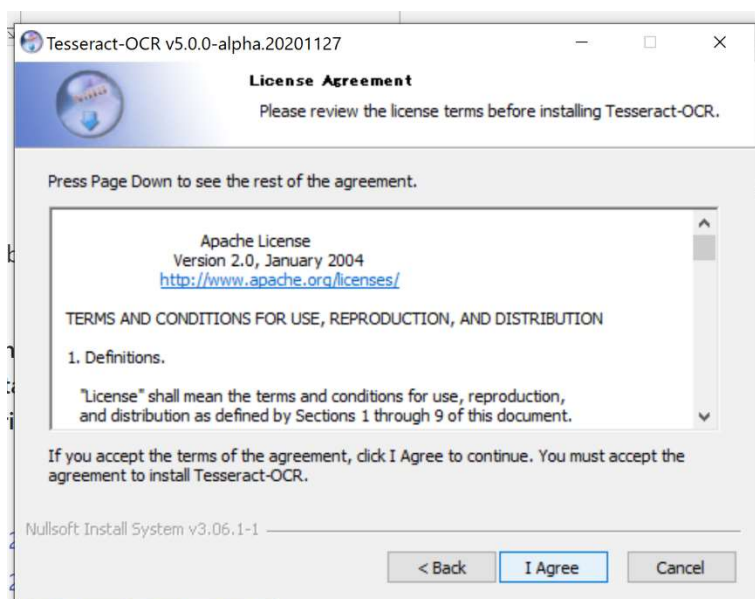
Clone this wiki locally

<https://github.com/UB-Mannheim/tesseract/wiki>

- ② Start the downloaded installer and select Yes / Next until the characters to be installed are selected (screen below).

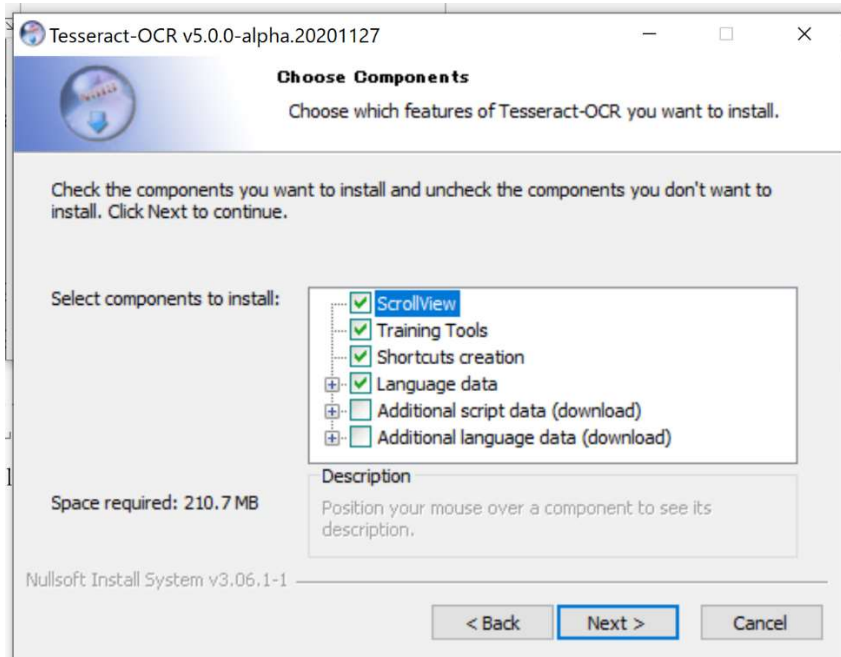


Next



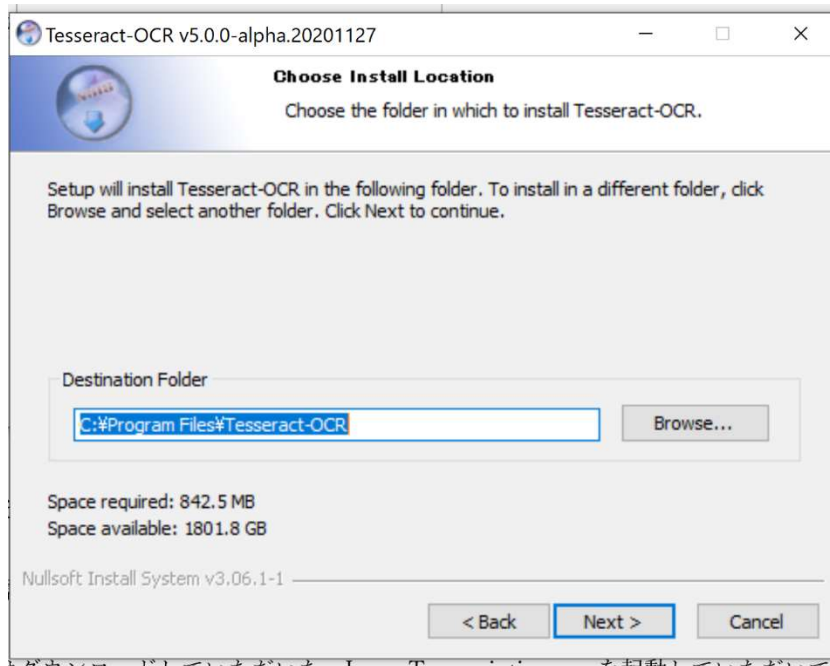
I Agree

- ③ Check all the language you need.

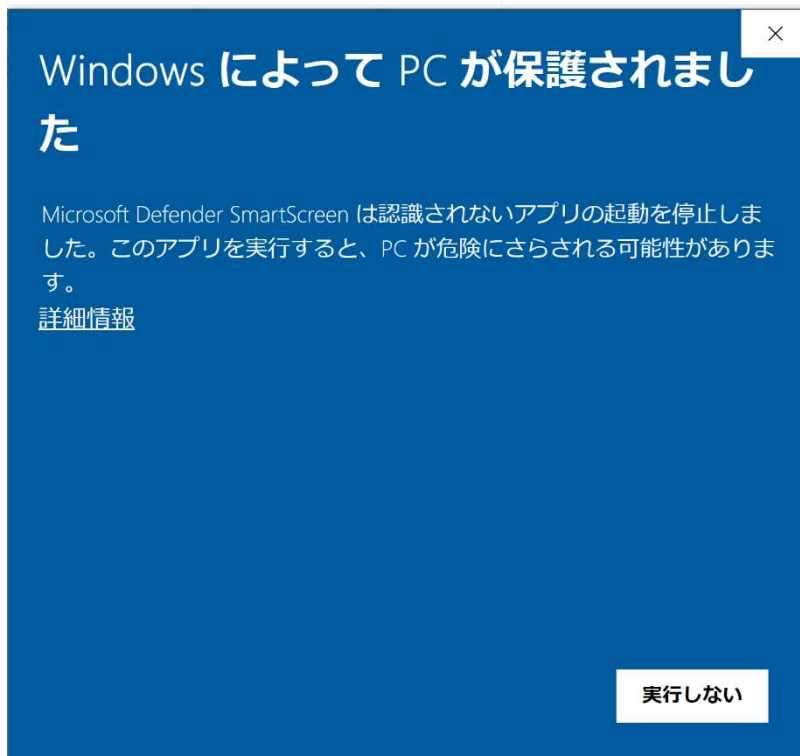


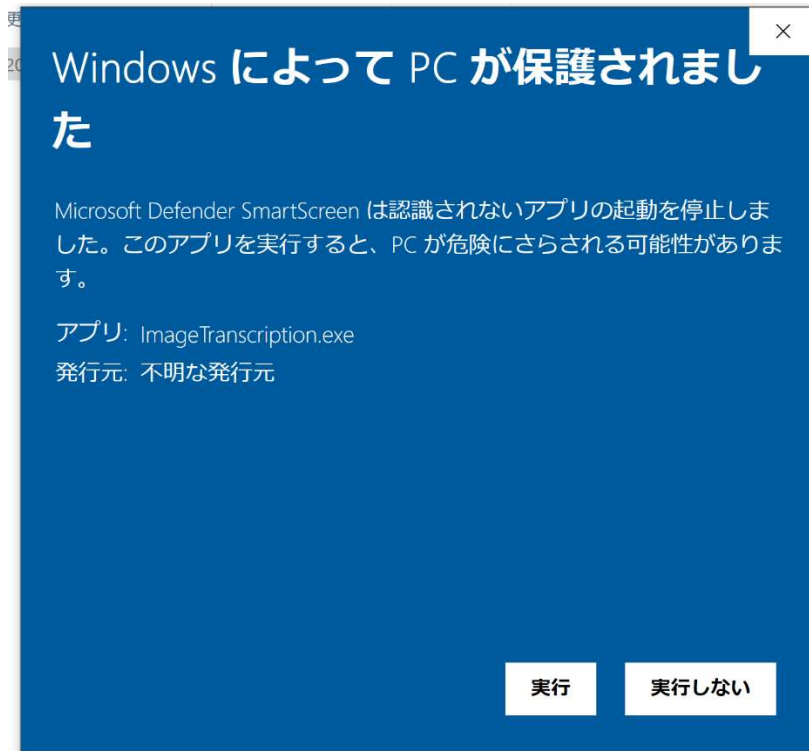
- ④ After that, you can click "Yes" or "Next".

Also, Note the folder name of the download destination.

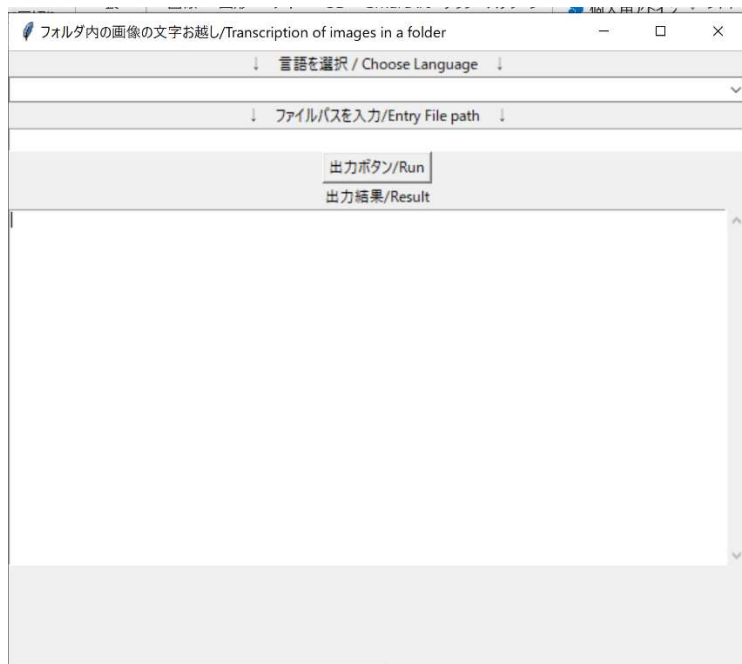


- ⑤ In the environment variable settings, add the Path you got in (4).
- ⑥ Please restart your Computer
- ⑦ Please download the transcript tool from our website.
- ⑧ After extracting the file, double-click the exe file in the folder.
- ⑨ Press the detailed information and then press "Run".

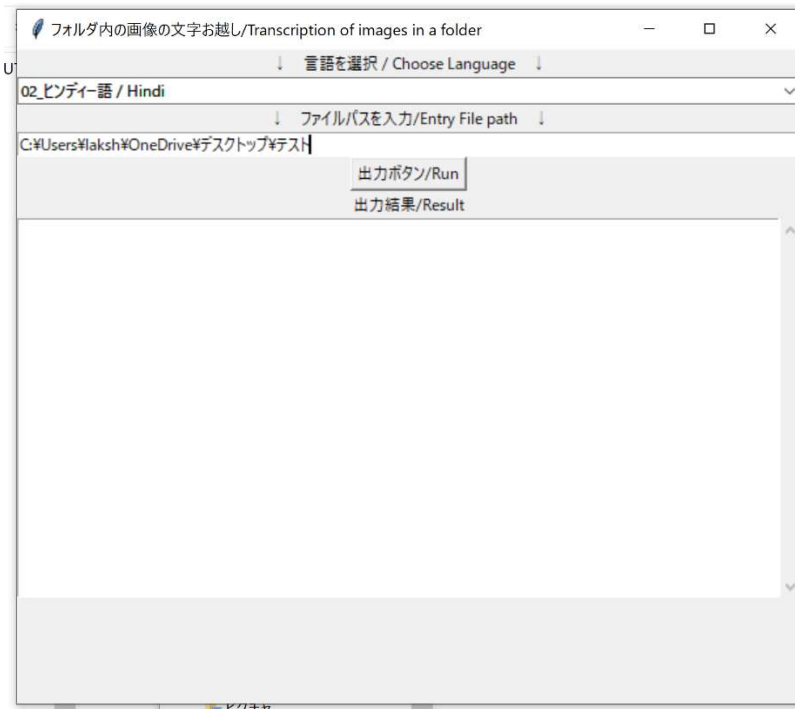




- ⑩ After a black screen like the command prompt appears, when the screen below starts up, you are ready to use.



- ⑪ Enter the language you want to read and the file path of the file containing the image you want to read, and press the "Output button".



- ⑫ It is output.
- ⑬ When using from the next time, follow the procedure of ⑩.

【important point】

- Specify the file path containing the image, not the file path of the image.
- Please note that if there is an image that does not need to be read in the specified file path, it will also be read.
- Please contact us freely if you have any question.