

OCR Extraction for Pre-Payment Processing

Installation Guide

Installation Steps*

1. Installing prerequisites: Python, Tesseract
 - a. Python
 - i. Install Python3 (version 3.6-3.7), following the instructions and download links found within the following link:
<https://www.python.org/download/releases/3.0/>
 - b. Tesseract
 - i. Install Tesseract-OCR, following the instructions found on the following link: <https://github.com/tesseract-ocr/tesseract/wiki>
2. Download the program files
 - a. Clone the repository from GitHub OR download as a ZIP and extract to desired directory.
 - i. When cloning, make sure to include the argument
`--recurse-submodules`
 - ii. Or, after cloning, `git submodule update --init --recursive`
3. Installing Python dependencies/requirements
 - a. Navigate to the top level directory of the project via Terminal/Command Line.
 - b. Run the command (Assuming you are using Python3. You may have to substitute pip for pip3): `pip3 install -r requirements.txt`
 - c. All required Python packages/libraries should now be installed.
4. Compiling TFWordBeamSearch.so for Handwriting Extraction functionality. If you correctly cloned the submodules, the CTCWordBeamSearch project should be located in the “external” directory.
 - a. Navigate to the top level directory of the project via Terminal/Command Line.
 - b. Run the command: `./compile_libs.sh`
 - c. The command from (b) should have compiled the library and copied it into
`src/main/backend/data_extraction/handwriting_extract/src/`
5. Installation has been completed!

Troubleshooting

- Error: Could not find a version that satisfies the requirement tensorflow
 - If you run into this error during step 3b, ensure that you have a version of Python between 3.0 and 3.7. As of writing this document, tensorflow does not support Python 3.8.

- Error: TFWordBeamSearch.so: undefined symbol: _ZN10tensorflow12OpDefBuilder4AttrESs
 - Should this error occur during step 4c, TFWordBeamSearch.so, open `external/CTCWordBeamSearch/cpp/proj/buildTF.sh` and remove all instances of `"-D_GLIBCXX_USE_CXX11_ABI=0"`, then rerun `buildTF.sh`.
 - If you have multiple versions of Python installed (i.e. `python3` vs. `python3.6`), you might have to adjust the file `external/CTCWordBeamSearch/cpp/proj/buildTF.sh`. Inside this file there are three conditionals that depend on the versions TensorFlow that you are using. Determine the version you have, and edit the two lines for `TF_CFLAGS` and `TF_LFLAGS` to use your version of python.
 - For example, you could change:
 - `TF_CFLAGS=($(python3 -c 'import tensorflow as tf; print("".join(tf.sysconfig.get_compile_flags()))')`
 - To be:
 - `TF_CFLAGS=($(python3.6 -c 'import tensorflow as tf; print("".join(tf.sysconfig.get_compile_flags()))')`

Verification

1. No errors are present.
2. Run `src/main/main.py`.
3. The file should run with no errors.
4. Navigate to the website (ie. `127.0.0.1:5000` if hosting locally).
5. The website should load with no issues.

*** The following installation guide assumes the usage of a Linux machine/server. The software has been configured and tested on Linux, and no guarantees can be made about installation on other operating systems.**