VIM OCR CONFIG

**JULY 2022  
  
  
ZEUS Environment Installation**

**Version 1.0.0.0**

By Milan .L. Hazra

OCR INSTALLATION GUIDE

Contents

[1 Prerequisites 3](#_Toc110031178)

[2 Python Installation and Environment Setup 4](#_Toc110031179)

[2.1 Download the Install Python 4](#_Toc110031180)

[2.2 Setting up Environment 7](#_Toc110031181)

[2.3 Installation of libraries 9](#_Toc110031182)

[3 Major External Libraries 11](#_Toc110031183)

[3.1 Tesseract-OCR Installation 11](#_Toc110031184)

[3.2 OCRMYPDF 16](#_Toc110031185)

[3.3 DB Browser SQLite installation 19](#_Toc110031186)

|  |
| --- |
| 1. Prerequisites |

|  |  |
| --- | --- |
| **System Requirements** | **Versions** |
| **Environment Version** | 1.0.0.0 |
| **OS** | Windows 10 |
| **RAM** | 8 GB |
| **GPU** | NVIDIA CUDA |
| **GPU MEMORY** | 2 GB |
| **Disk Space** | 1 GB |
| **Python** | 3.9.2 |
| **OCRMYPDF** | 13.6.2 |
| **Tesseract-OCR** | 5.2.0 |

|  |
| --- |
| 1. Python Installation and Environment Setup |

* 1. Download the Install Python
* Go to the following link - <https://www.python.org/downloads/>.

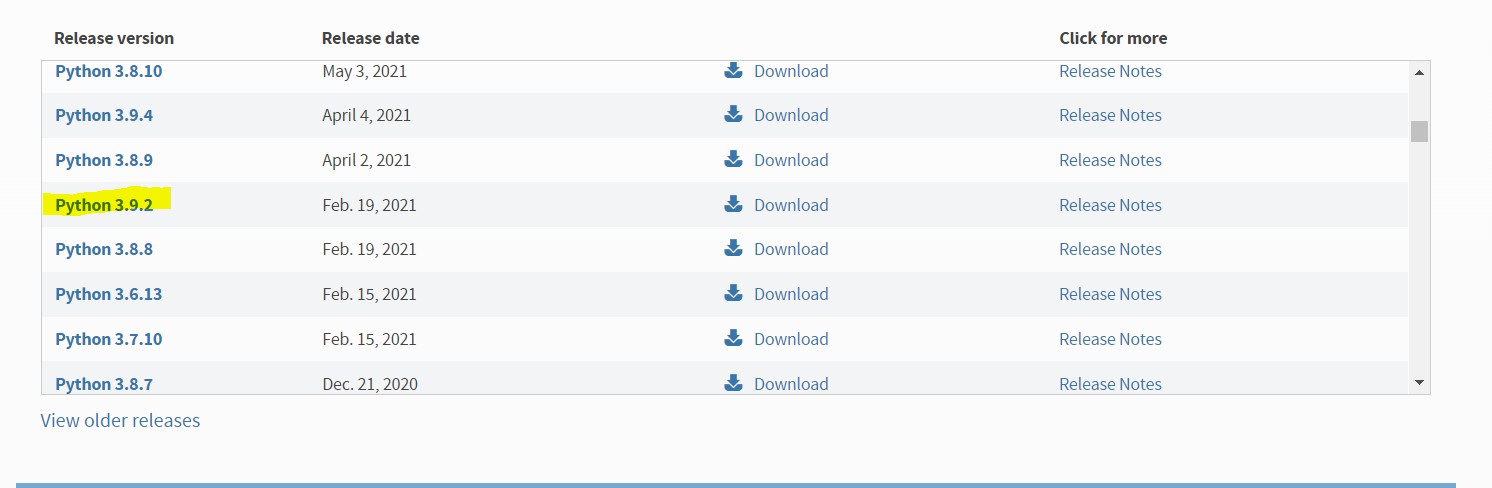


Figure 1

* Click on Download and scroll Down Click on the link marked .

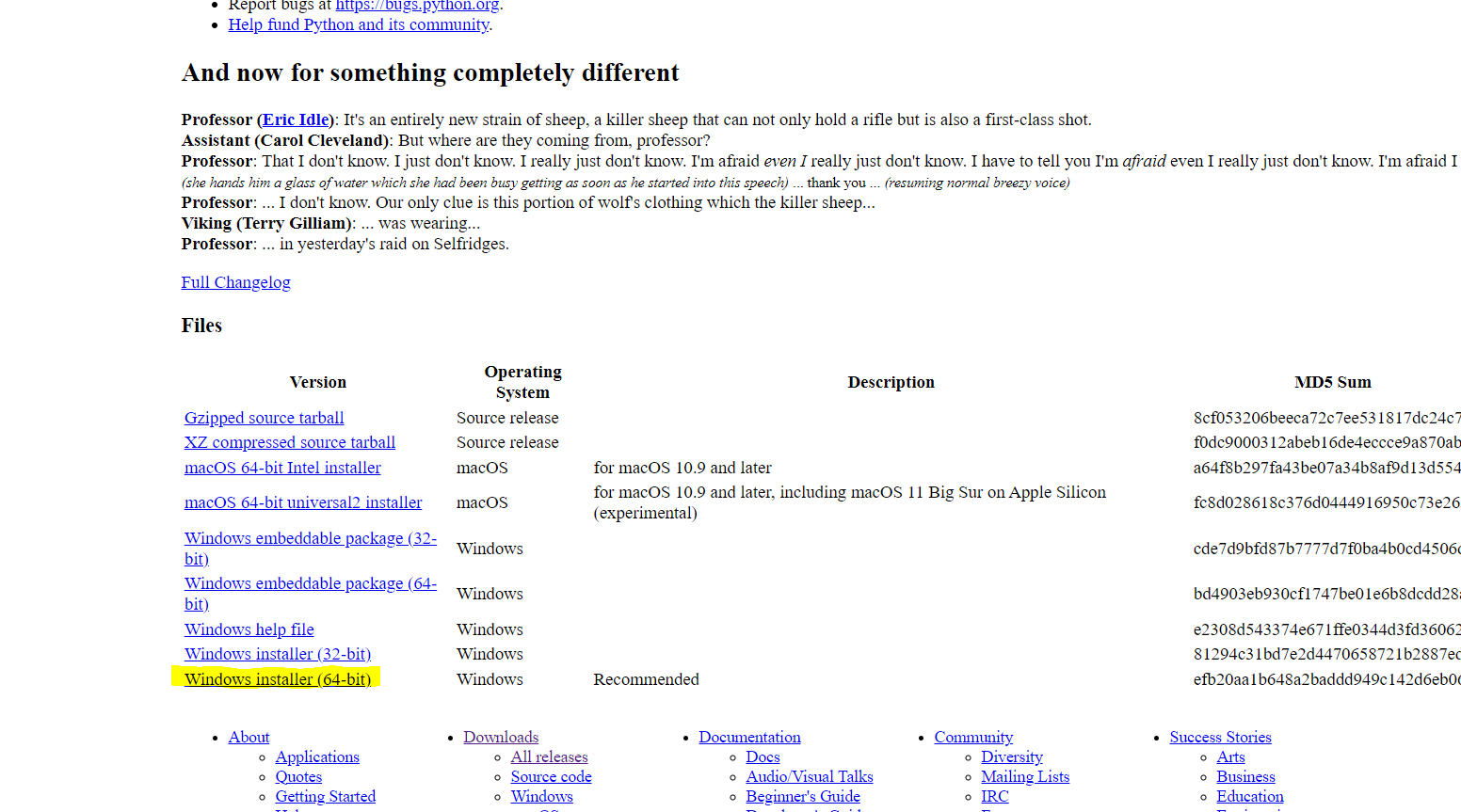


Figure 2

* You will see Windows installer of Python is donloaded.

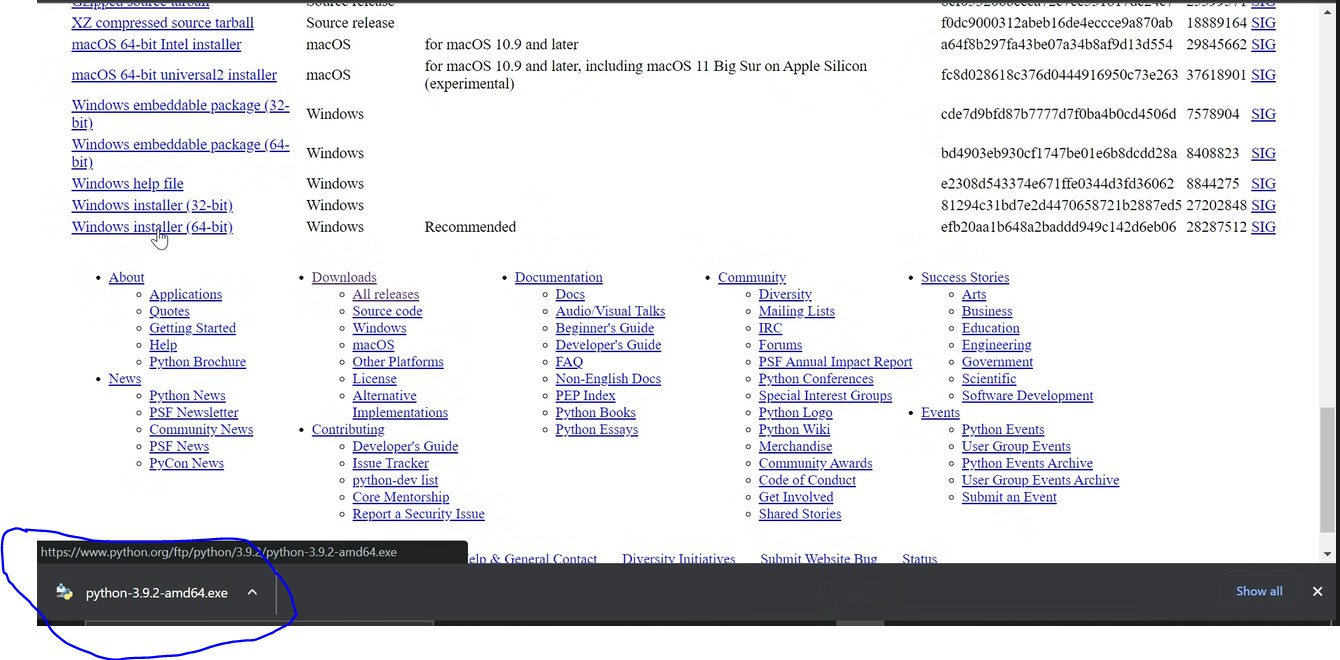


Figure 3

* After Clicking on .exe file you will see the Python Installer Screen. Please **tick** the option of

**ADD PYTHON 3.9 to PATH**

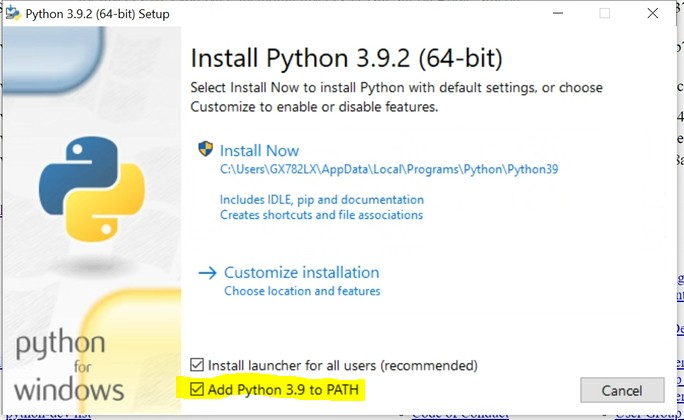


Figure 4

* Click on Install Now

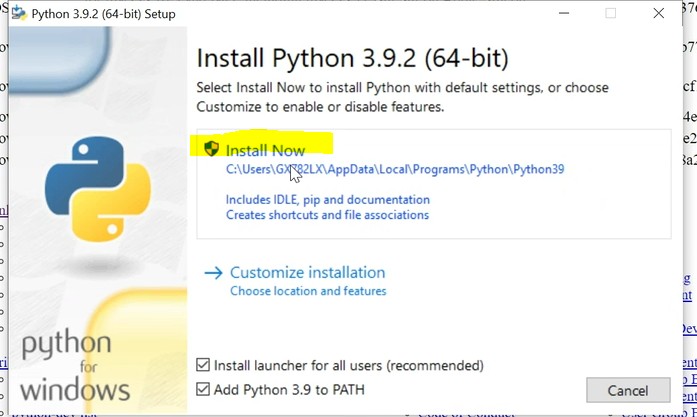


Figure 5

* Once the installation started you will see the setup progress screen

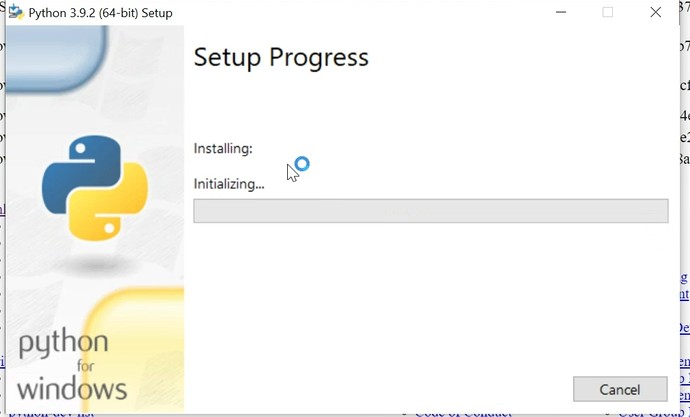


Figure 6

* Once the installation is done you will find the below messages on your screen

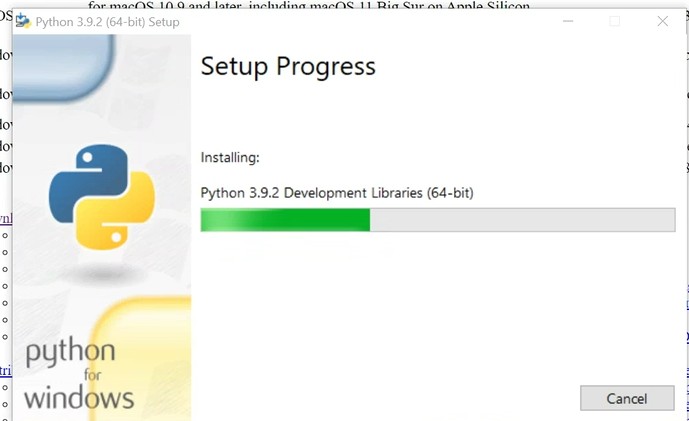


Figure 7

* After Successful installation Go to Windows Search Bar and type CMD and check if python is installed . To check type the following command.   
  python

After typing the following command if python shell is displayed then the python is installed properly.

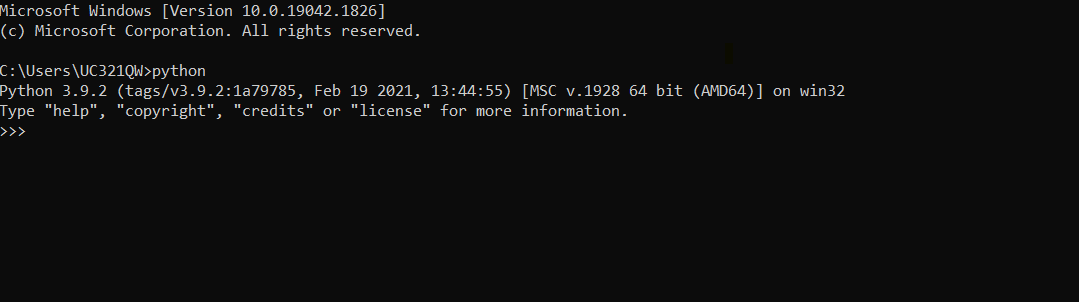


Figure 8

* 1. Setting up Environment
* Using the **exit()** command in the shell, come out of the shell
* Navigate to C folder using following commands.

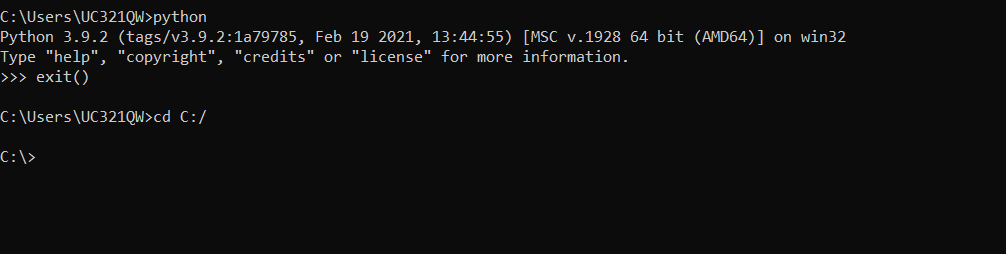


Figure 9

* First we have to install Virtual Environment package and then create virtual environment.

Run the following command to install virtual environment.

**pip install virtualenv**

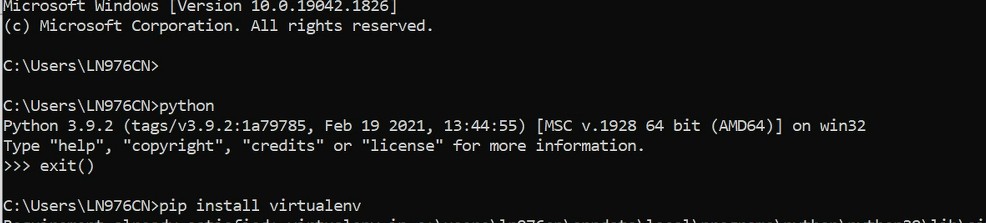


Figure 10

* Check if Virtualenv is installed or not using the following command.

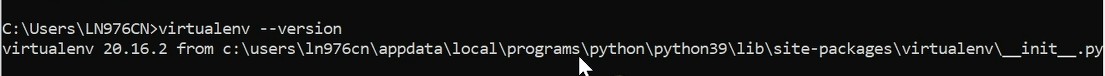


Figure 11

* Create a folder in C Drive with desired name and navigate to the folder.

Run the following command to create a virtual Environment.

* **virtualenv <name\_of\_environment>**  
  Note – below given example ‘**zeus**’ is the name of the environment.

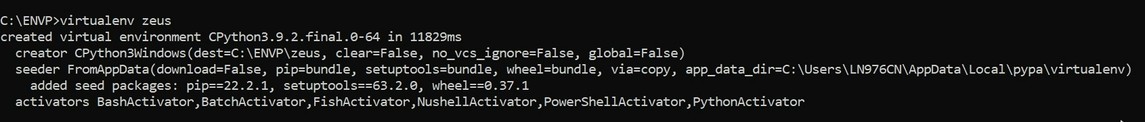


Figure 12

* Once the environment is created using the above command you will see that a folder is created in your project folder with the name of the environment you have provided.

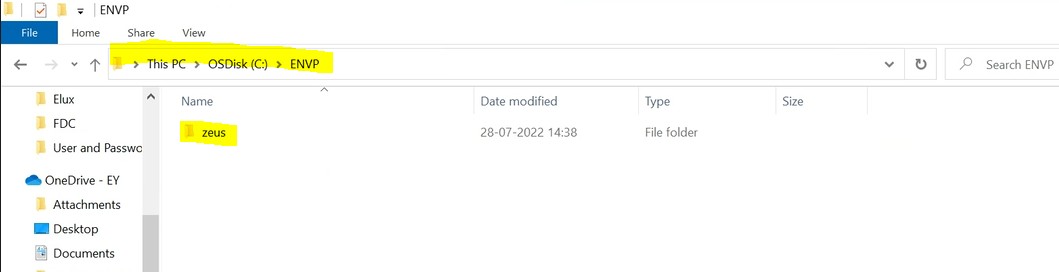


Figure 1

* For activating the environment use the following command by navigating to your project folder in environment is saved.

**<name\_of\_environment>\Scripts\activate**

After activating you can see that before your navigation script in CMD (zeus) has been created which means that the environment is activated now.

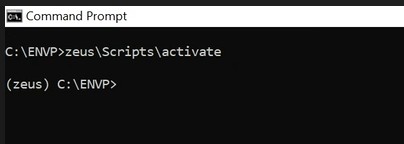


Figure 14

* To deactivate the environment use the following command

**<name\_of\_environment>\Scripts\deactivate**

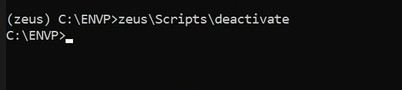


Figure 15

**You have successfully set up your environment. Next step will be installing the libraries inside the environment.**

**The main reason for creating the environment is that it will be easy to export libraries and we can work on different project without disturbing the configuration of other projects due to installation of libraries especially in python. Therefor it is always recommended to use environment for a particular project.**

* 1. Installation of libraries
* Download and save the below requirements.txt file in your Project Folder



* Once the requirements.txt file is downloaded run the below command
* **pip install -r requirements.txt**

After running the command you will see list of libraries getting installed

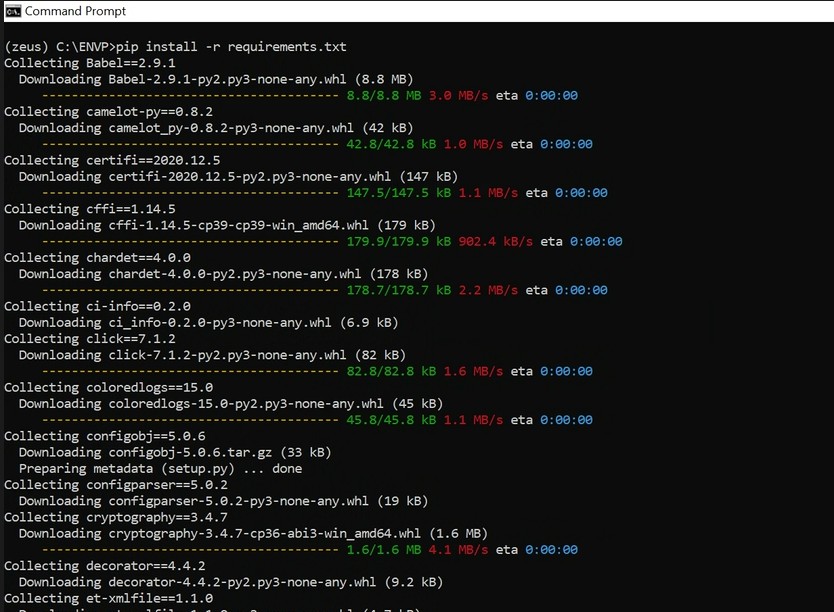


Figure 16

* Once all the libraries are installed successfully you will see all the libraries just like in the below screenshot.

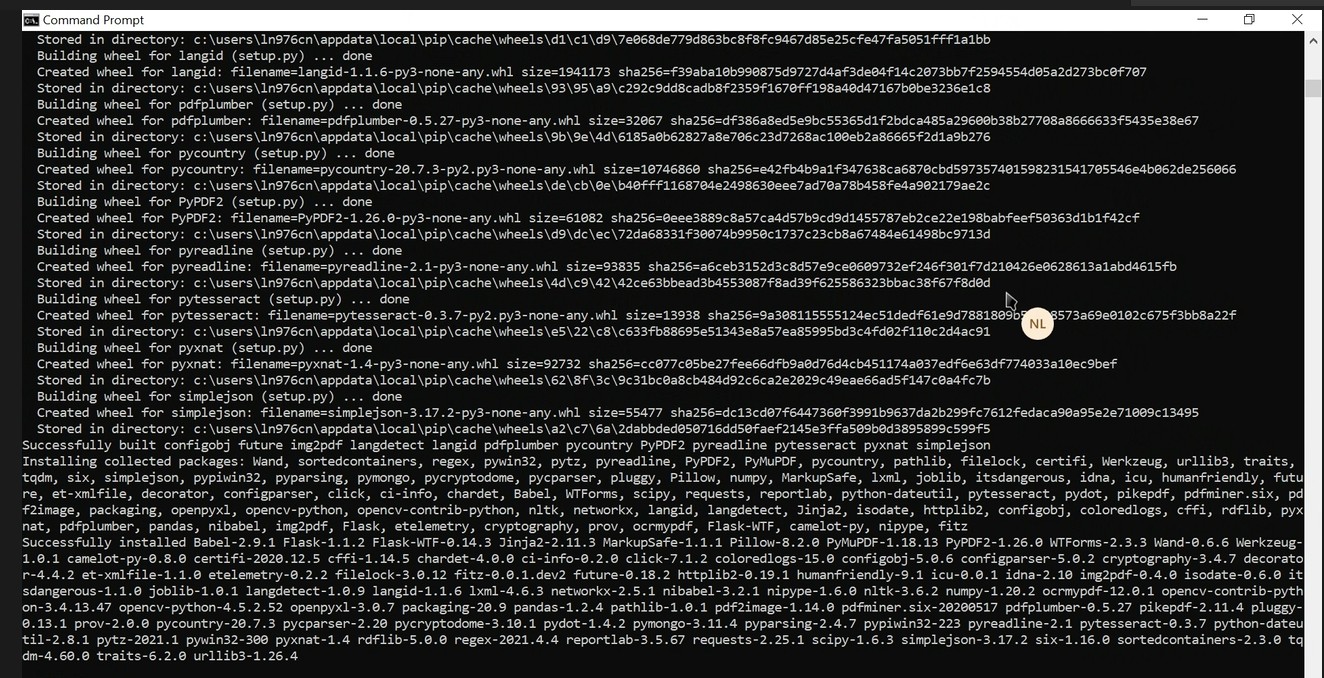


Figure 17

|  |
| --- |
| 1. Major External Libraries |

* 1. Tesseract-OCR Installation
* Tesseract installation is one of the important and challenging part please follow the steps closely.
* Go to the following link - <https://github.com/UB-Mannheim/tesseract/wiki>

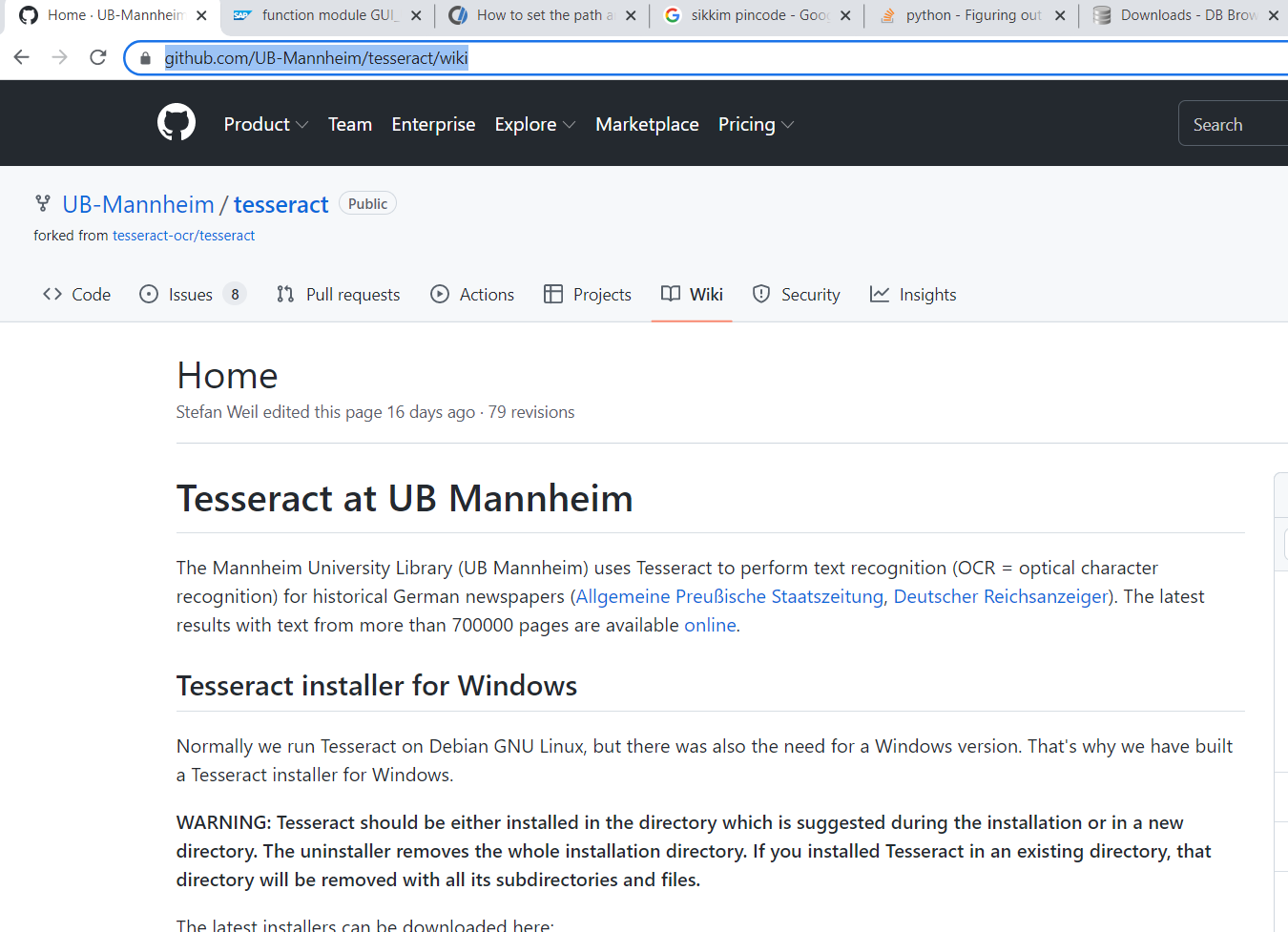


Figure 18

* Scroll Down and click on the following link with exe.

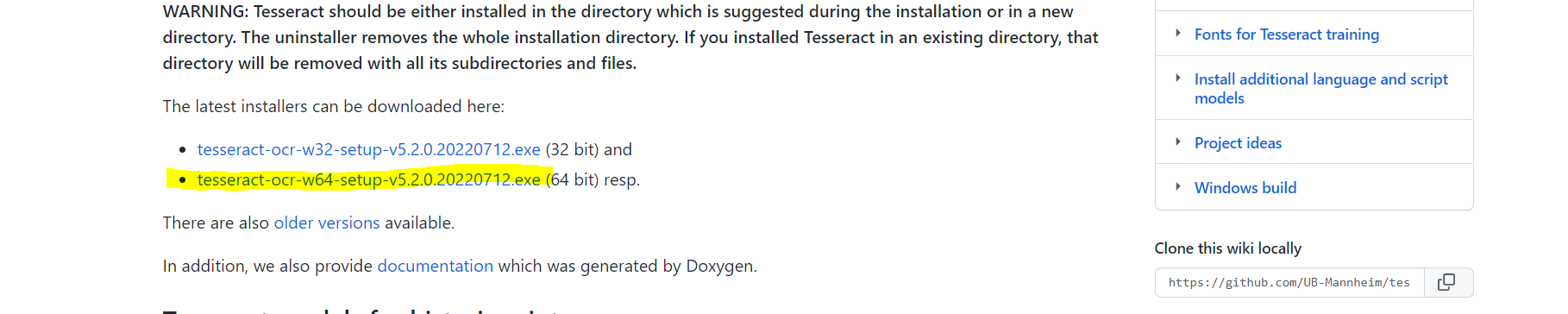


Figure 19

* Click on the downloaded exe .

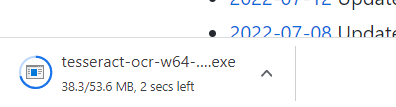


Figure 20

* Once the exe is clicked you will see the installation in progress screen.



Figure 21

* Once the installation is done you will find the following success message .

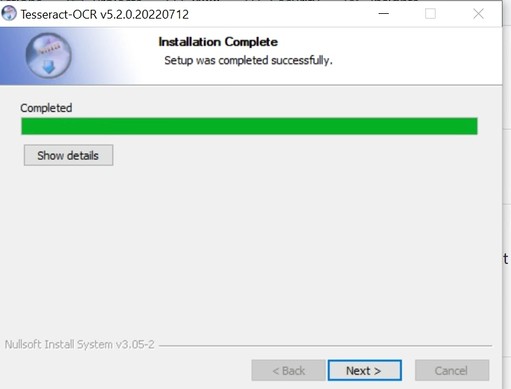
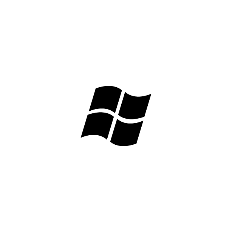


Figure 22

* Press  . **+ R** key and type **%AppData% and press ENTER key.**

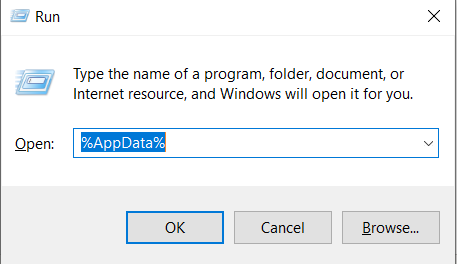


Figure 23

* You will reach the following folder. Once you reach the folder click on **AppData** in the Navigation Bar.

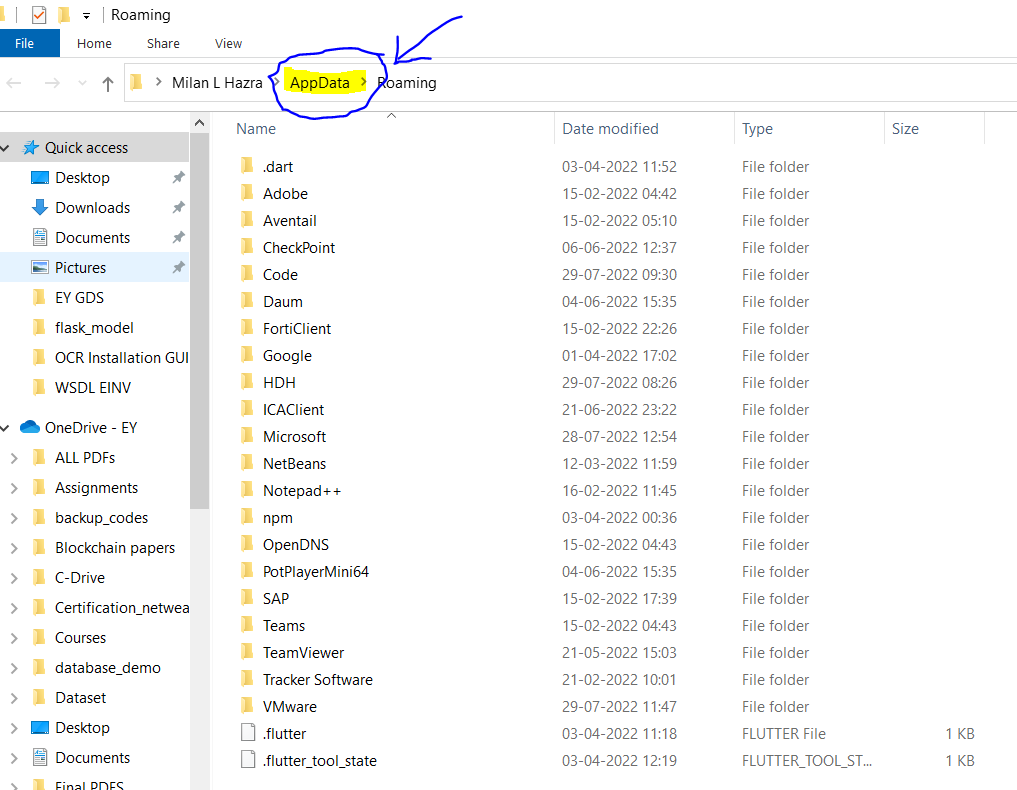


Figure 24

* Once you click the AppData you will see following following folders. Go to Local folder and check for folder named **Tesseract-OCR.**

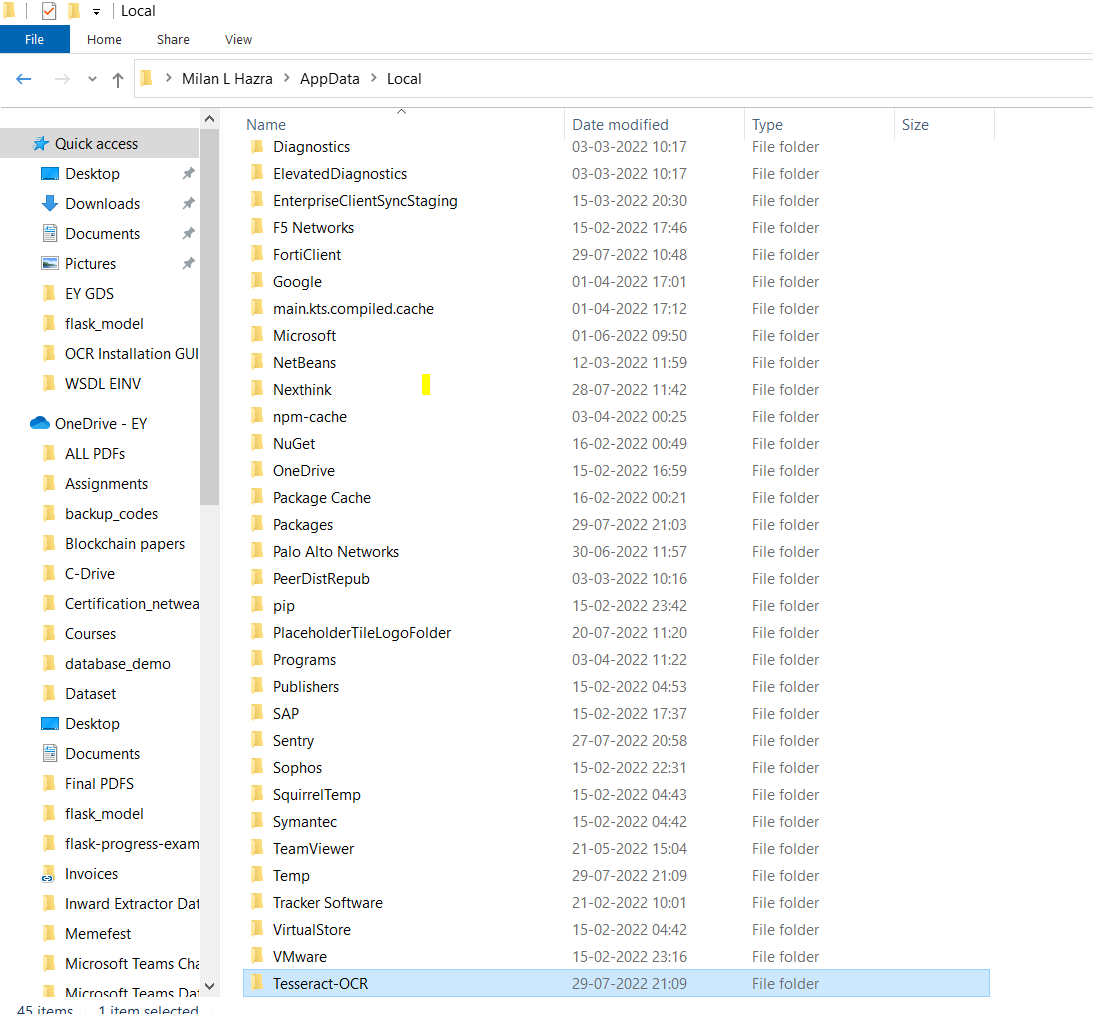


Figure 25

* Copy the following folder and paste in **C:/Program Files.**
* **Delete the file from /AppData/Local/ folder afterwards.**
* **Go to Environment variables**

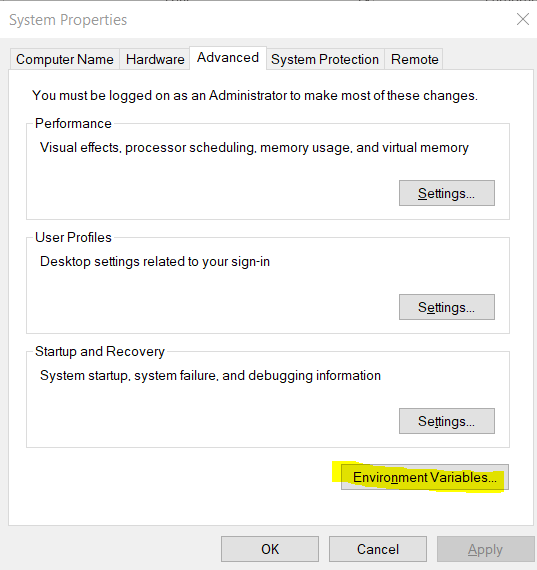


Figure 26

* **Maintain the path in Environment Variables**.

**C:\Program Files\Tesseract-OCR**

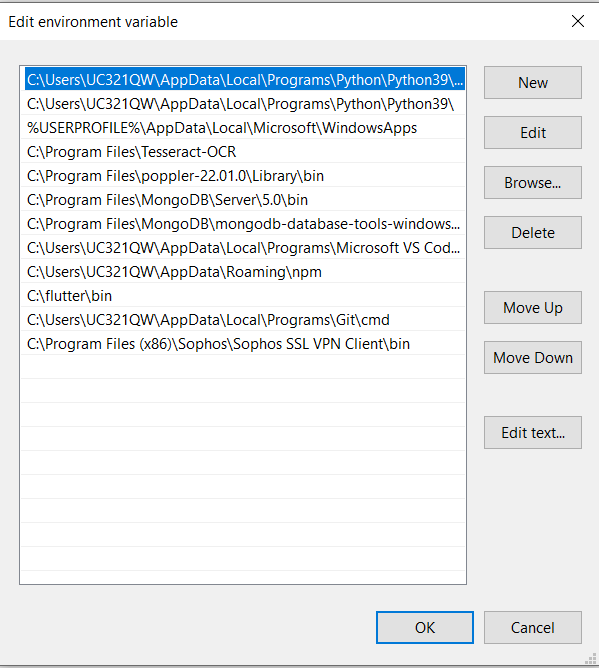


Figure 27

* Once the path is save Go to Command Prompt and type the following command to check if the OCR is installed successfully
* Command **- tesseract**

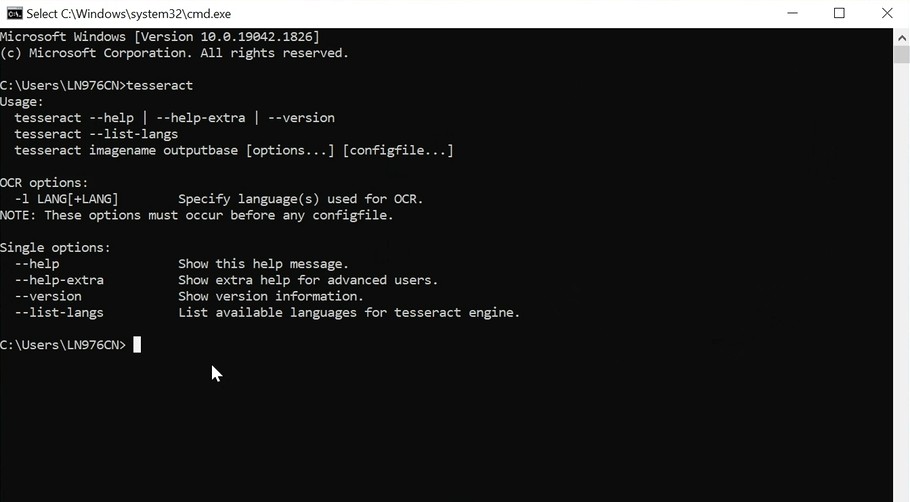


Figure 28

* **Tesseract OCR is successfully installed** 
  1. OCRMYPDF
* **OCRMYPDF** this particular library is essential to handle the image based and text based pdf configuration .
* Please follow the steps very closely.
* For installation OCRMYPDF you must install the following for Windows:

Python 3.7 (64-bit) or later(already installed in above process)

Tesseract 4.0 or later(already explained in above process)

Ghostscript 9.50 or later

* For Ghostcript installation please follow the process given below.
* First we need to install Chocolatey package manager install the following using powershell.
* Open your powershell in administrator mode.

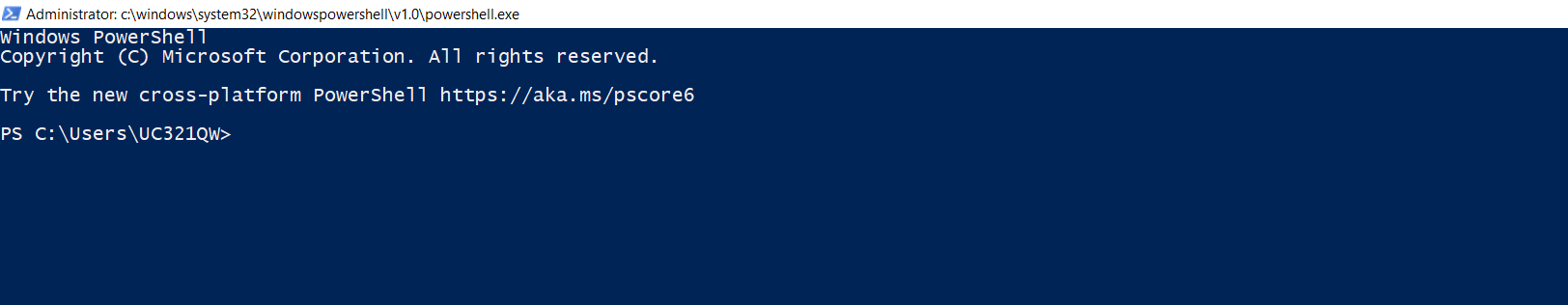


Figure 29

* Run the command given below in the Windows Powershell.

Command - **Set-ExecutionPolicy Bypass -Scope Process -Force; [System.Net.ServicePointManager]::SecurityProtocol = [System.Net.ServicePointManager]::SecurityProtocol -bor 3072; iex ((New-Object System.Net.WebClient).DownloadString('https://community.chocolatey.org/install.ps1'))**

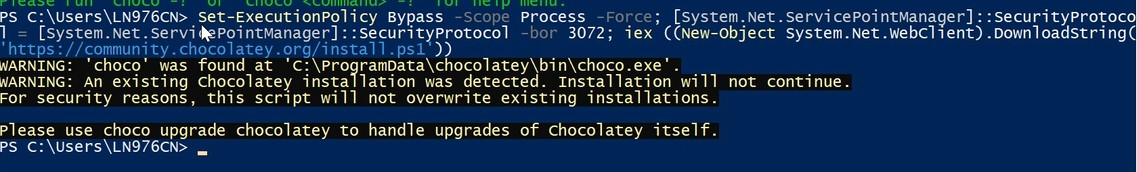


Figure 30

* After the commands run successfully check if the chocolatey package is installed by typing the following command.

Command – **choco.**

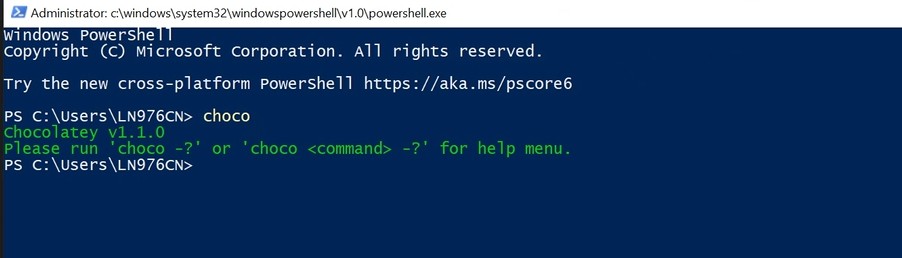


Figure 31

* If version appears then it is installed successfully.
* After the chocolatey package installation successfully we need to install ghostscript using the following command.

Command – **choco install ghostcript**

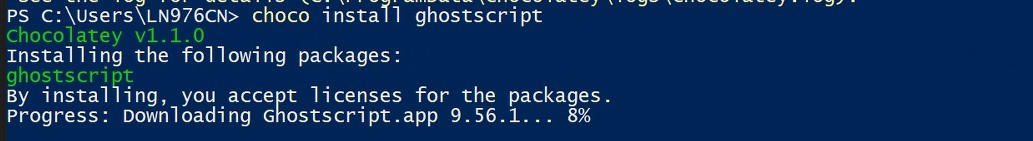


Figure 32

* If version appears then it is installed successfully.
* After installing ghostscript we can install OCRMYPDF using the following command

Command – **pip install ocrmypdf**

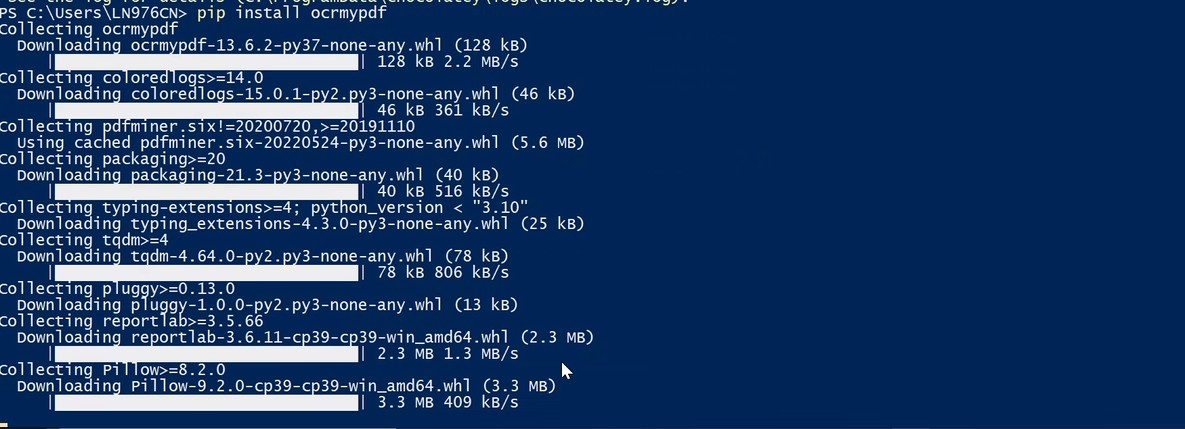


Figure 33

* After installation you can close the powershell and open command prompt .
* Using command prompt go to the project folder using command **cd C:/ENVP**
* And activate the environment using command **zeus\Scripts\activate.**
* Once you are inside the environment type the following command

Command – **pip install ocrmypdf==13.6.2**

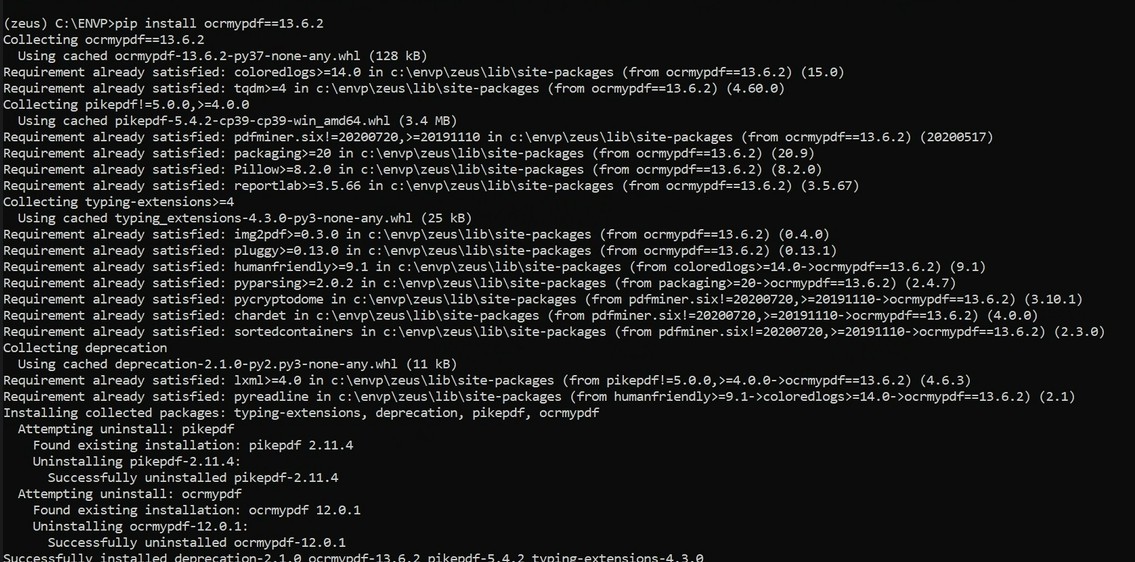


Figure 34

After that check if ocrmypdf is installed successfully in this environment using the following command – **ocrmypdf –version**



Figure 35

**OCRMYPDF is installed successfully**

* 1. DB Browser SQLite installation
* Go to the following link - <https://sqlitebrowser.org/dl/>
* Download the link marked in the image with yellow highlighter .

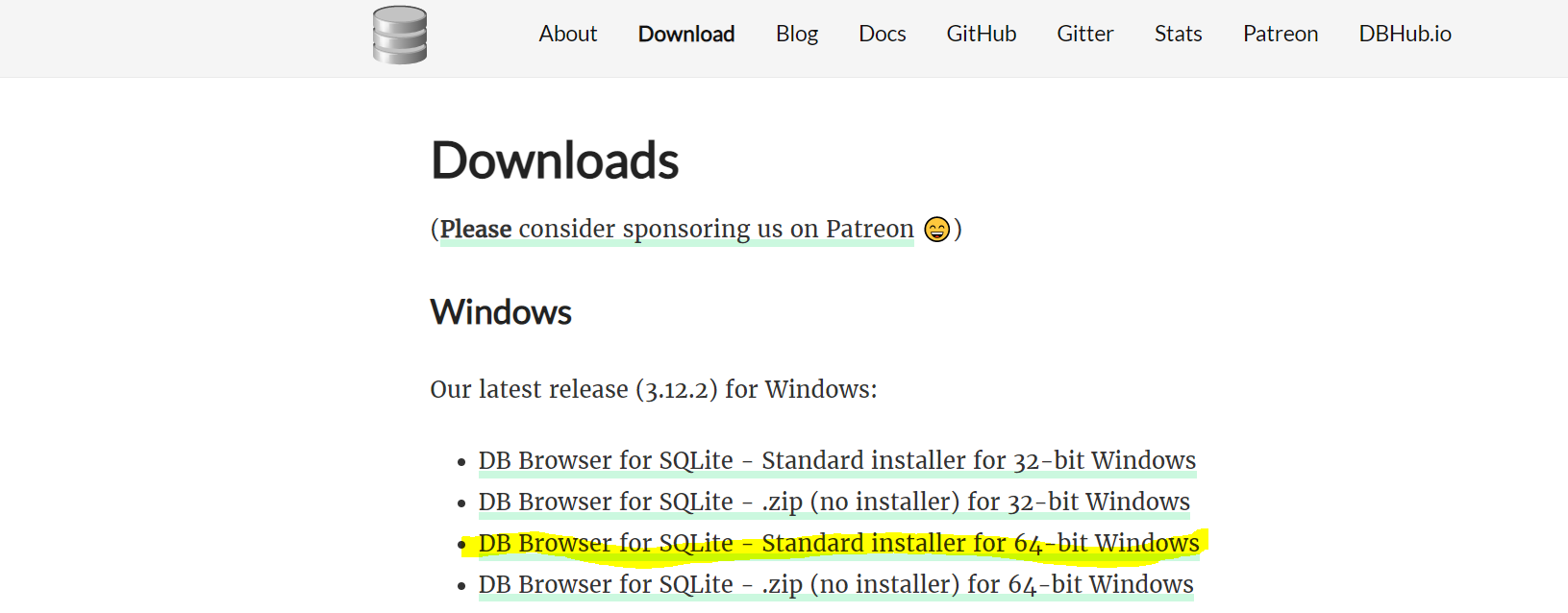


Figure 36

* Click on the link and download the installer. Click on the installer and you will get the following screen.

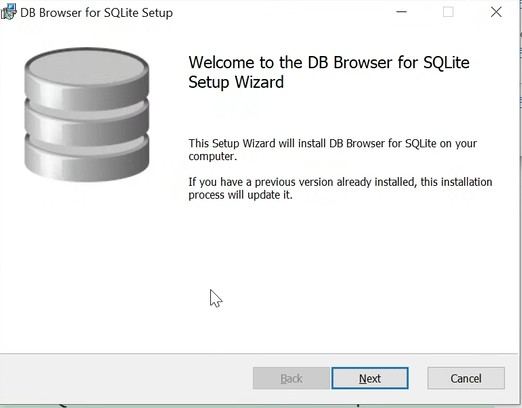


Figure 37

* Click on Next and you will see the following screen of *installation in progress* .

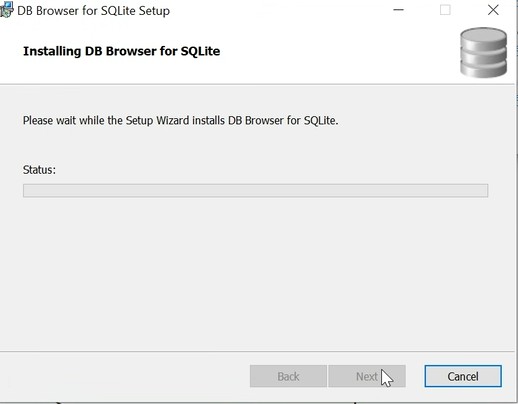


Figure 38

* After the installation is complete, we can open **DB Browser for SQLite .**

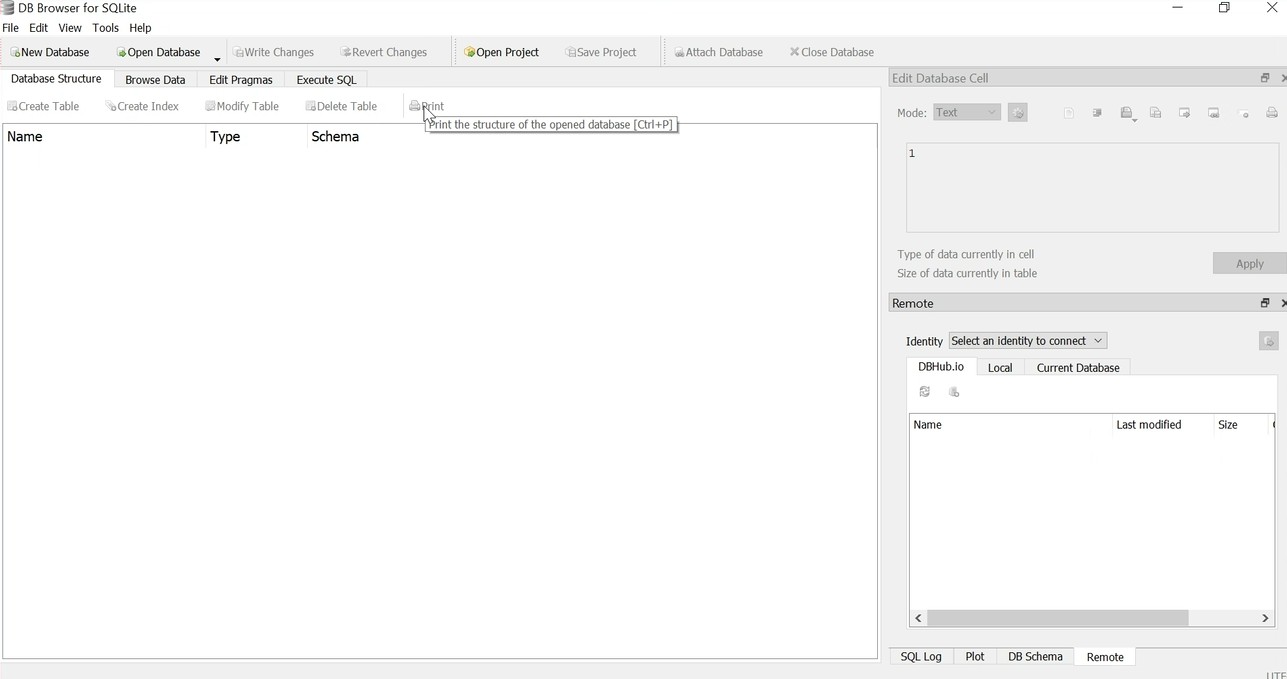


Figure 39

**Installation of all the libraries is finished.**