PDF to CSV on Python: Setup and Installation

NOTES:
→ PLEASE only use pure Ubuntu or Mac when using Camelot, NOT Windows and NOT WSL, it is not worth the pain
→ A combination of pip, conda, and regular sudo installs are used in this guide, this is the combination that worked with zero issues in setup and so is recommended but other install methods *should* work

Anaconda (+Python) install (also installs most required libraries) -Skip steps already done previously

→Update system repositories
$ sudo apt update

→Download curl
$ sudo apt install curl -y

→ Enter temp directory
$ cd /tmp

→ Head to https://repo.anaconda.com/archive/ to check that the below download link (https://repo.anaconda…) matches the most recent version, if it does not replace with latest

→Download Anaconda installer script
$ curl --output anaconda.sh
https://repo.anaconda.com/archive/Anaconda3-5.3.1-Linux-x86_64.sh

→ Run installer script
$ bash anaconda.sh

→ press **ENTER**

→ scroll down to bottom of user agreement and type '**yes**'

→ press **ENTER** to confirm location (should be something like /home/anaconda3)

→ wait couple of minutes for installation

→ activate environment settings
$ source ~/.bashrc

→ check install (4.12.0 in June 2023)
$ conda --version

→check conda is working and that you have the required libraries
$ conda list
→ check for ‘numpy’, ‘pandas’, and ‘ghostscript’
→ if any of these libraries are not already installed run the appropriate commands and recheck afterwards
$ sudo apt -y install ghostscript
$ pip install pandas
$ pip install numpy

→ (optional) create directory for jupyter files (good practice as makes navigating inside jupyter just simpler)
$ mkdir jupyter_folder
→enter directory
$ cd jupyter_folder

→ open jupyter
$ jupyter notebook

→ congrats! Jupyter should now be open on a local host in your browser and you can create new Jupyter notebooks to write and run Python scripts!

Camelot

⟶should have already checked for ghostscript and so just need to check for tkinter
$ python -m tkinter
→ window with tkinter interface should appear on screen, if not install
$ sudo apt install python3-tk
→ rerun check to test installation

→ install camelot
$ conda install -c conda-forge camelot-py

→check installation
$ camelot --version

→ That is it! You can now proceed to the code itself:
https://github.com/mwinterdata/pdf_to_csv/blob/main/PDF%20to%20Excel%20Code.ipynb