1) Download Tesseract OCR Engine from https://github.com/UB-Mannheim/tesseract/wiki

   Then run the exe file and install it, make sure it is in C:\Program Files\Tesseract OCR

   Then go to “Environment Variables” to “System Variables” in “Path” add a new with the path ‘C:\Program Files\Tesseract OCR’

   Sources:
   https://stackoverflow.com/questions/46140485/tesseract-installation-in-windows (First answer)

   https://linuxhint.com/install-tesseract-windows/

2) Download poppler and put it in your environment variables.
   https://github.com/oschwartz10612/poppler-windows/releases

   Follow the instruction in the first answer here
   https://stackoverflow.com/questions/18381713/how-to-install-poppler-on-windows

3) Download the .zip file of github repo
   https://github.com/impira/docquery

   Extract it in the same folder of the notebook you’re using for installations

   Take the folder named “docquery-main” and make sure it is
   In the notebook folder

   In notebook run

```
!cd docquery-main && pip install .[all]
```

4) Install other libraries

```
!pip install transformers==4.23
!pip install pydantic==1.10.8
!pip install pymupdf
!pip install spacy
!pip install sentencepiece
!pip install pytesseract
!pip install numpy==1.20.0
```

5) Run the notebook of V2.0 “the latest”