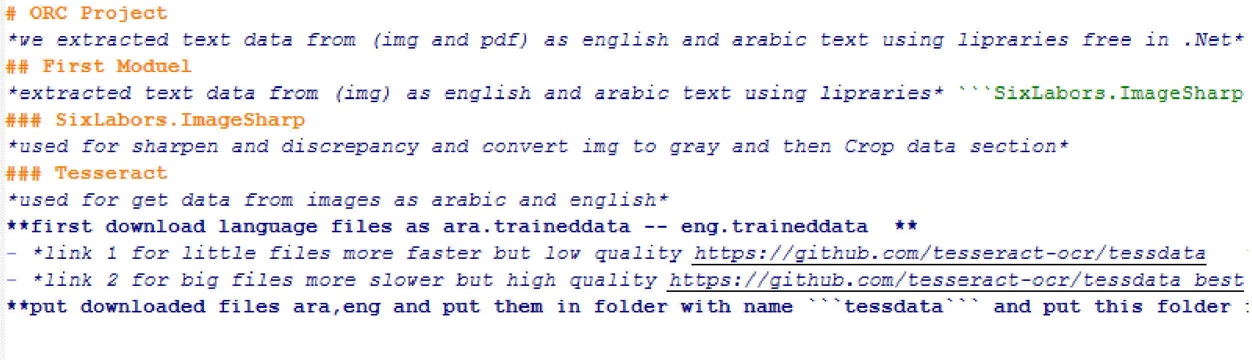
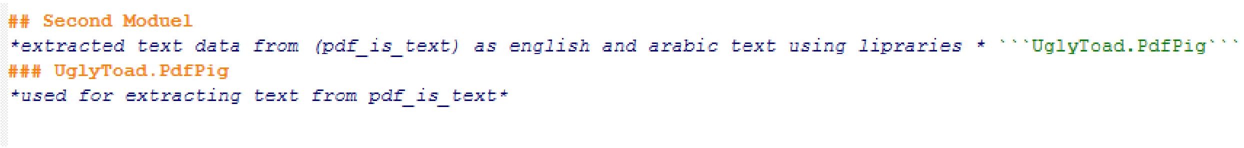
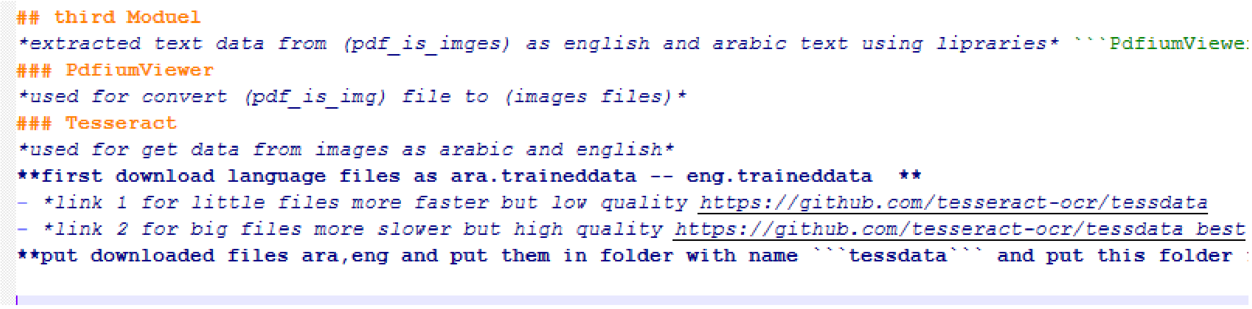


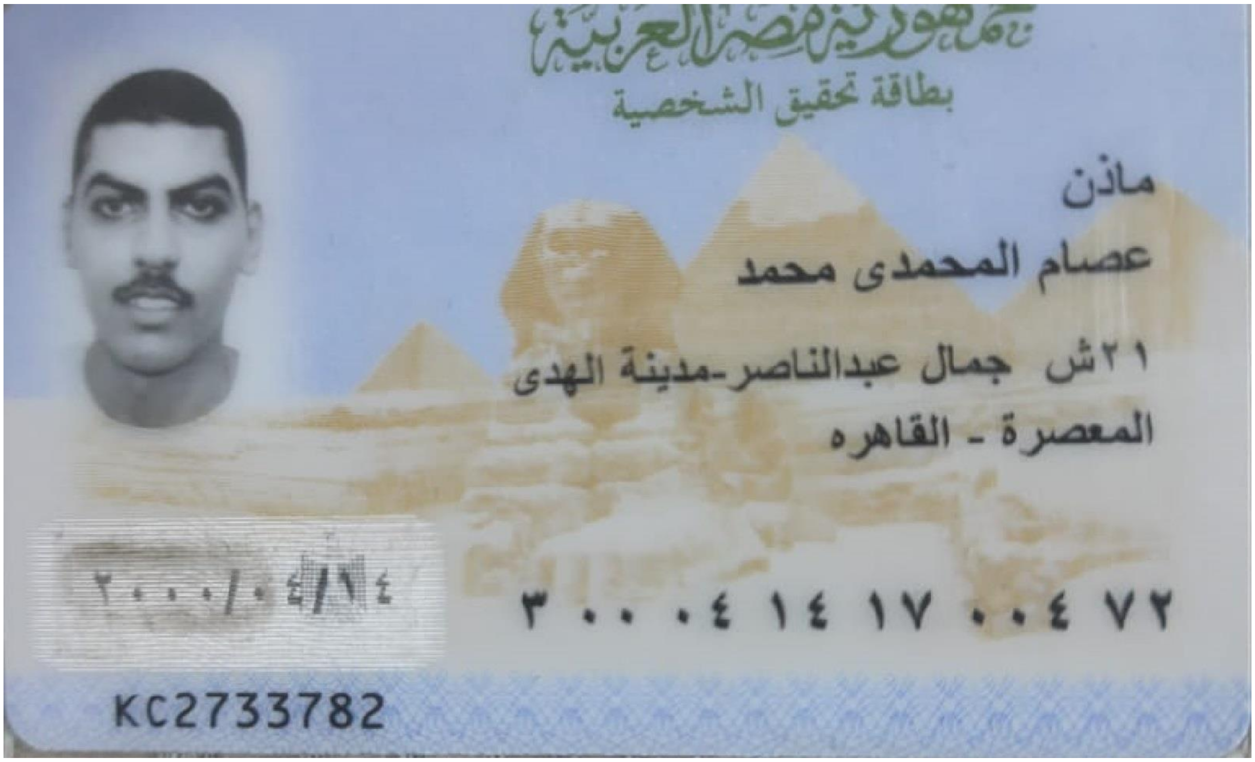
```
# ORC Project
*we extracted text data from (img and pdf) as english and arabic text using lipraries free in .Net*
## First Moduel
*extracted text data from (img) as english and arabic text using lipraries* ```SixLabors.ImageSharp
### SixLabors.ImageSharp
*used for sharpen and discrepancy and convert img to gray and then Crop data section*
### Tesseract
*used for get data from images as arabic and english*
**first download language files as ara.traineddata -- eng.traineddata  **
- *link 1 for little files more faster but low quality https://github.com/tesseract-ocr/tessdata
- *link 2 for big files more slower but high quality https://github.com/tesseract-ocr/tessdata_best
**put downloaded files ara,eng and put them in folder with name ```tessdata``` and put this folder
```



```
## Second Moduel
*extracted text data from (pdf_is_text) as english and arabic text using lipraries * ```UglyToad.PdfPig```
### UglyToad.PdfPig
*used for extracting text from pdf_is_text*
```



```
## third Moduel
*extracted text data from (pdf_is_imges) as english and arabic text using lipraries* ```PdfiumViewe
### PdfiumViewer
*used for convert (pdf_is_img) file to (images files)*
### Tesseract
*used for get data from images as arabic and english*
**first download language files as ara.traineddata -- eng.traineddata  **
- *link 1 for little files more faster but low quality https://github.com/tesseract-ocr/tessdata
- *link 2 for big files more slower but high quality https://github.com/tesseract-ocr/tessdata_best
**put downloaded files ara,eng and put them in folder with name ```tessdata``` and put this folder
```



جمهورية مصر العربية

بطاقة تحقيق الشخصية

ماذن

عصام المحمدى محمد

٢١ش جمال عبدالناصر-مدينة الهدى
المعصرة - القاهره

٢٠٠٠/٠٤/١٤

٣ ٠٠ ٠٤ ١٤ ١٧ ٠٠٤ ٧٢

KC2733782