All important links during learning of ocr

- use word as developer to create form
- this guy made s/w in .net can ask for help
- As pdf don't work convert it with specify density, depth
- its sublink are useful specially hocr
- Preprocessing, binarize
- everything u want to know
- Java application for same
- improve efficiency pdf

For single character recognization
tesseract $image $outbase -psm 10   You need to set Tesseract's page segmentation mode to "single character."

For empty page err: leave 10-12 pixels above or below image.
Or add option -psm 10

How to improve OCR accuracy

1. https://docparser.com/blog/improve-ocr-accuracy/#more-994
2. **Use tesseract version 4.0**
3. **Hocr, pypdfocr, unpaper**

Form processing part

- **Don't use absolute coordinates, instead use relative (Typically, x/y would be a percentage of width/height instead of an absolute pair of values)**

Just Type :- how to train tesseract for handwritten text

- The number of fonts is limited to 64 fonts.
- Note that runtime is heavily dependent on the number of fonts provided, and training more than 32 will result in a significant slow-down.

1 method
Take advantage of handwritten fonts.

2 method

Create your own font.

Miscellaneous techniques for form processing

- Histogram can be used for contrast, thresholding….etc. pixel intensity vs pixelCount()
-

For a regular sized font of about 11pt a good resolution is about 300 to 500 DPI here it is.