<https://www.pyimagesearch.com/2017/07/10/using-tesseract-ocr-python/>

<https://thebittheories.com/from-scanned-pdfs-to-text-summarisation-part-1-6a34e2485911>