

2022-11-7

• # A new method to blog

~~Paper website~~

Paper Website is a service that lets you build a website with just pen and paper. I am going to try and replicate the process. ||

&#182;

## The Plan

The continuity feature on macOS + iOS lets you scan PDFs directly from your iPhone. I want to be able to scan these pages and automatically run an Automator script that takes the PDF and OCR the text. Then I can further ~~can~~ clean the text and convert from markdown. ||

&#182;

## ## Challenges

The OCR software I planned on using cannot detect my shitty handwriting. ||

I can try to write a bit neater or use something like Azure ||

&#182;

## ## Solution

### ### OCR

In the PDFs, all the scans are saved as images on a page. I extract the image and then send it to Azure's API ||

&#182;

### ### Fixing Syntactical errors

Using OpenAI's model as described in the original post is a great way to fix errors. ||

&#182;

### ### Paragraph Breaks

The recognized text has multiple line breaks. Therefore, I use what is called a pilcrow to specify paragraph breaks. ||

~~###~~ ¶;

### ## Where is the code?

I created a [GitHub Gist] for a sample Python script to take the PDF and print the text

¶;

A more complete version, including some hacks is on my [GitHub] and [Gitea]

¶;

\*\* In Part 2, I will discuss some more features \*\*