Zackary Morelli

1/14/2022

[zackmorelli@gmail.com](mailto:zackmorelli@gmail.com)

PatientScansOCRBackgroundWorker is a windows service that I made. As in a service that runs in the background of the OS. It monitors the Aria documents folder for any changes that are made for when a new scanned patient document is added, as a PDF. When this happens, the PatientScansOCRBackgroundWorker windows service starts the Tesseract Execute program as a separate process. This would run as a service on my computer, but by running the service under my user account, it was able to monitor the Aria file server, which is a network location of course, because I have access to it with my user account.

The Tesseract Execute program uses the Tesseract OCR library to read in text saved as images in scans of patient documents that are saved to the Aria documents folder. It was simply easier to have the service run Tesseract Execute on a separate process than try to run Tesseract in the service itself, just because a windows service is different than a normal executable program. The Tesseract Execute program would save the output of the Tesseract OCR as a simple text file that could be read by the Plancheck Eclipse script.

Of course, all of this would ideally have been done in Plancheck itself, but I had to come up with this big workaround because I was not able to install the .NET C++ runtime files needed for utilizing the Tesseract library (which is a C# wrapper for a C++ program) on the Aria file server. So this workaround allowed me to run it locally, and have the text files already there when someone ran Plancheck, with the document check module.

This service successfully ran in the background on my computer for around a month, before it was made obsolete by commercial software.

The Tesseract OCR library is good, but keep in mind it still depends on the quality of the scanned document.