

Klippa engineering challenge

I hereby share the engineering challenge with you. You have at maximum 1 week to complete it. You can share a private github repo (or similar) with us when you are done so we can review it. Please keep in mind that we review all aspects of coding. So documentation, variable naming, coding conventions, performance and of course functionality.

The assignment:

Create a CLI tool that allows you to call our OCR API.

The CLI tool must be able to do the following:

- Provide the API key as an option
- Provide the template as an option
- Provide PDF text extraction fast or full as an option.
- Provide a file that it processes via the OCR API (PDF or image).
- Display the result of the processing in a nicely formatted way.
- Include the option to save the json output to the file as {file name}.json

Bonus points for:

- Being able to process a folder instead of 1 file and being able to batch process the entire folder
- For making it run in Docker
- Keep track of the totals of the folder processing, such as VAT percentages and total amount and display that after processing
- Monitor a folder, so when a new file is added to the folder it will be processed automatically.
- Even more bonus points if you manage to do the above things concurrently, so if you process a whole folder you process several at once
- Provide options for API key, template and PDF text extraction through a global config file (e.g. like Docker and Git: ~/.docker/config.json and ~/.gitconfig)

The documentation for the OCR API can be found here: <https://custom-ocr.klippa.com/docs>

We have created an API key for you, with template financial_full. It has 100 credits:
JNkH5eZwuVCDTYK2t5gbrOpJXXSfy1MG