F.A.Q

Dúvidas frequentes

Como adiciono um diretório além do CGTI para realização do OCR?

Crie o diretório em questão - recomenda-se seguir a estrutura atual, ou seja, `/srv/shares/novoSetor/ocr/` - e **crie** o arquivo .watchmanconfig dentro de “entrada” com o texto `{"settle": 10000}`. Como também **crie** os diretórios "saída" e "originais". Após estas etapas, **rode o comando** de criação das triggers: Pronto, novo diretório para processamento OCR adicionado.

Como testar o OCR para um diretório que eu adicionei?

Apenas **copie/envie** o PDF para o servidor, **fora do diretório .../ocr/entrada/**. Então, logado no servidor via linha de comando, **copie** o PDF para dentro do diretório .../ocr/entrada que você criou anteriormente. O PDF será processado normalmente.

Existe algum log da solução OCR?

O arquivo de LOG presente em `/var/log/ocr.log` indica o estado de cada arquivo quanto ao seu processamento.

Como sei que o OCR está instalado?

Rode o comando: `ocrmypdf --version`

Deve aparecer um resultado similiar a: `5.7.0`

Mensagens de comando não encontrado indicarão um problema na sua instalação, portanto, revise todas as etapas antes de continuar.

Como sei que o PDF está sendo processado no servidor?

O último estado no LOG para o arquivo em questão deve ser "Processando". Além disso, por meio do comando `top`, você deverá ver linha(s) com a execução de `gs`, `tesseract` ou `ocrmypdf`. Isto indica que um ou mais arquivos estão sendo processados.

Por que o processamento demora tanto em arquivos grandes?

Quanto maior o número de páginas ou características do documento escaneado, maior o tempo necessário para o processamento.

Existe alguma maneira de acelerar o processamento do PDF?

Incrementando a quantidade de cores e memória principal da sua VM é possível processar os arquivos PDFs em menor tempo.

O arquivo está recebido pela solução OCR mas não está sendo processado, o que pode estar acontecendo?

O watchman pode não estar ativo. Caso isso ocorra, rode o comando `watchman watch-list` e observe se os processamentos foram iniciados no log da solução OCR. Caso não haja entradas no log, verifique o log do watchman:

```
$ tail -f /usr/local/var/run/watchman/root-state/log.
```

Mensagens de erro podem ser consultadas no Google para dúvidas.

Caso o watchman esteja OK e ainda haja problemas para processar o arquivo, verifique se existe disco e memória suficientes para realizar esta operação.

Ao mover um arquivo PDF de dentro do próprio servidor, o OCR funciona. Mas enviando do meu usuário via rede (samba, sftp, outro meio) o OCR não acontece, o que posso verificar?

Problemas de permissão local. Siga as políticas definidas para o acesso aos diretórios e certifique-se de que ao enviar o PDF via rede, as permissões estejam adequadas e ajustadas para atender a sua política adotada no campus.

Adicionalmente, alguns softwares de envio de arquivos, ou mesmos os servidores de arquivos, possuem mecanismos para ajustar as permissões padrão para novos arquivos enviados.

Revisão #4
Criado 2 April 2018 14:30:35 por Aleciano Ferreira
Atualizado 20 August 2020 00:55:35 por Aleciano Ferreira