**DOCUMENTAÇÃO**

Linguagem: Python

Bibliotecas: Docx2txt

Métodos:

* Percorrer pasta raiz e gravar em uma lista, o nome dos arquivos da pasta
* Criar arquivo vazio ‘paginasContadas.txt’
* Percorrer lista de nomes e caso for ‘.docx’, converter usando docx2txt, contar quantidade de caracteres – numero de quebras de linhas ‘\n’, adicionar contagem ao total, escrever no arquivo ‘paginasContadas.txt’

Procedimentos;

V1:

* Primeiros testes em arquivos simples, contagem correta com o word
* Após teste com arquivos mais complexos, houve divergência entre as contagens.
* Após isolar casos, foi constatado que há contagem dupla de caracteres em caixas de textos, uma solução será estudada.

V2:

* Após uma analise na biblioteca docx2txt, foi desenvolvido um método para eliminar linhas duplicadas no texto, que funcionou. Porém, ao analisar mais fundo o método aplicado, viu-se que ele eliminava linhas importantes do texto e não apenas duplicadas das caixas de texto. Outro método de abordagem será estudado.

V3: