

# **Documentação teste 2**

**Autor:**

Antonio César

# Conteúdo

<b>1</b>	<b>Funcionamento</b>	<b>2</b>
<b>2</b>	<b>Versões</b>	<b>2</b>
<b>3</b>	<b>Bibliotecas utilizadas</b>	<b>2</b>
3.1	zipfile . . . . .	2
3.2	urllib.request . . . . .	2
3.3	tabula . . . . .	2
3.4	os . . . . .	2
<b>4</b>	<b>Funções</b>	<b>2</b>
4.1	request.urlretrieve . . . . .	2
4.2	ZipFile . . . . .	3
4.3	convert_into . . . . .	3
4.4	write . . . . .	3
<b>5</b>	<b>Problemas</b>	<b>3</b>

# 1 Funcionamento

Baixa o arquivo `anexo1.pdf` da página do governo brasileiro, converte todas as tabelas do arquivo em CSV e compacta o resultado em `Teste_Cesar.zip`.

## 2 Versões

- OpenJDK 11.0.15;
- javac 11.0.15;
- python 3.8.10;
- Ubuntu 20.04 86\_x64;
- VSCode 1.67.2.

## 3 Bibliotecas utilizadas

### 3.1 zipfile

Usada para criar o arquivo zipado.

### 3.2 urllib.request

Usada para baixar o arquivo requisitado a partir de uma url.

### 3.3 tabula

Usada para converter PDF para CSV.

### 3.4 os

Usada para remover os arquivos da cache depois de utilizados. É preciso baixar a biblioteca com o comando `'pip install tabula-py'`.

## 4 Funções

### 4.1 request.urlretrieve

Baixar um arquivo a partir de uma url. Parâmetros:

- url
- diretório onde o arquivo será salvo

## 4.2 ZipFile

Cria um arquivo ZIP. Parâmetros:

- caminho onde o arquivo será salvo
- permissões
- tipo de compressão
- nível de compressão

## 4.3 convert\_into

Converte um PDF para CSV. Parâmetros:

- caminho do PDF
- caminho onde o CSV será salvo
- tipo da saída
- páginas que serão convertidas

## 4.4 write

Escreve o CSV o arquivo ZIP. Parâmetros:

- caminho do CSV
- arquivo específico que será zippado

# 5 Problemas

A conversão ocorre com sucesso, porém o cabeçalho de todas as tabelas são passados para o CSV, tentei usar funções da biblioteca pandas para consertar, mas as funções só reconhecem 1 coluna.