



Script para reducción de archivos PDF en proceso de carga masiva

Introducción

Durante el proceso de carga masiva (ya sea de programas, proyectos, transferencias y/o transacciones), es necesario subir al sistema los archivos de respaldo. Estos archivos comúnmente son en formato PDF obtenidos electrónica o digitalmente. Sin embargo, muchas veces durante la creación de estos archivos PDF no se considera la reducción del peso, lo que es posible lograr sin afectar la calidad del contenido. Hacer esta reducción es útil en varios sentidos:

- Reducción del tamaño del archivo ZIP utilizado para el proceso de carga masiva
- Disminución en el tiempo utilizado para la carga del archivo ZIP.
- Disminución en el tiempo de procesamiento del archivo ZIP.
- Disminución en los tiempos de descarga de los archivos PDF en el momento de revisar las rendiciones
- Mejor aprovechamiento del repositorio documental del sistema y del sistema de respaldo.

Por este motivo, la Contraloría General de la República, CGR, ha desarrollado un script que permite realizar reducción de los archivos tipo pdf, sin sacrificar la calidad (resolución) de los mismos y de manera recursiva (no uno por uno), el que pone a disposición de todos los servicios públicos/privados que utilizan SISREC¹.

Requisitos

- Windows 10 64 bits o superior.
- Python 3.0 o superior (incluido en zip de descarga del script, en carpeta bin).

¹ Este script puede ser utilizado para reducir cualquier archivo PDF, no solo los de SISREC.



- La ruta a los binarios de Python deben estar configurado en la variable de entorno PATH. Esto lo hace el programa de instalación de Python en el primero paso (ver imagen del punto 3 del procedimiento de instalación, más adelante en este mismo documento).

Procedimiento

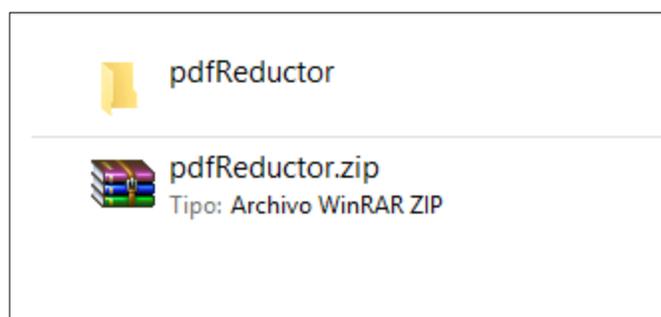
En resumen, el script se encuentra escrito en lenguaje Python² y utiliza un programa open-source llamado GhostScript³ (incluido en zip de descarga del script y no es necesaria su instalación para Windows) que permite el manejo de archivos PDF.

El script solicita al usuario una carpeta “raíz” en donde se encuentran todos los archivos PDF, y luego recorre esa carpeta y todas las subcarpetas interiores buscando archivos PDF y reduciéndolos.

Finalmente, y al mismo nivel de la carpeta raíz solicitada, se genera una nueva carpeta con el mismo nombre pero con el posfijo *_reduced*, en donde estarán los archivos PDF reducidos.

El procedimiento para el sistema operativo Windows es el siguiente:

1. Descargar el archivo ZIP de instalación del script desde www.rendicioncuentas.cl desde la sección Documentos y archivos de interés.
2. Descomprimir el archivo ZIP descargado en alguna carpeta a su elección, generando una carpeta llamada *pdfReductor*



3. Dentro de la carpeta *pdfReductor*, hay una carpeta llamada *bin\pythonwin* y que contiene el instalador del lenguaje Python para Windows y proceder a su

2 <https://www.python.org/>

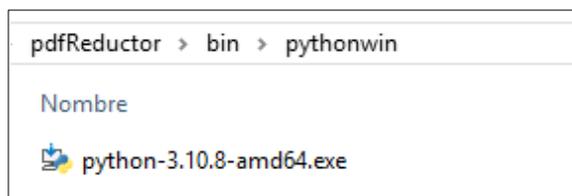
3 <https://www.ghostscript.com/>



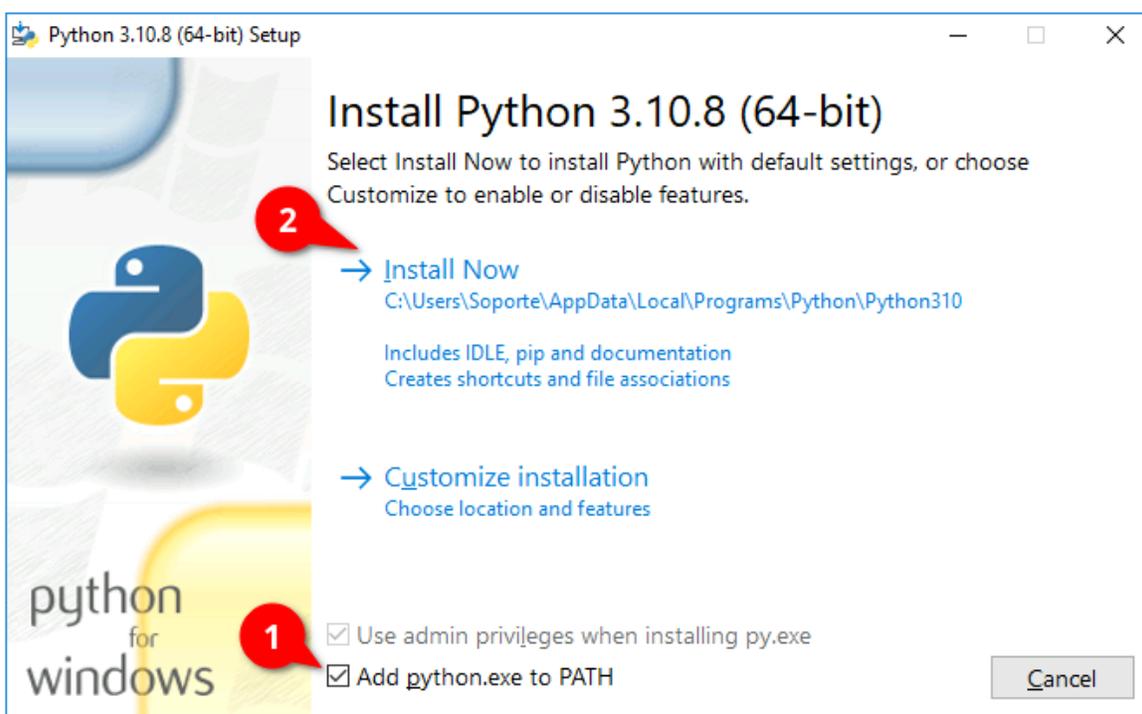
UNIDAD DE RENDICIÓN ELECTRÓNICA DE CUENTAS



instalación.



Durante el proceso, el instalador le preguntará si desea agregar Python al PATH de Windows, lo que debe ser confirmado.



Una vez, terminada la instalación de Python, es necesario reiniciar Windows.

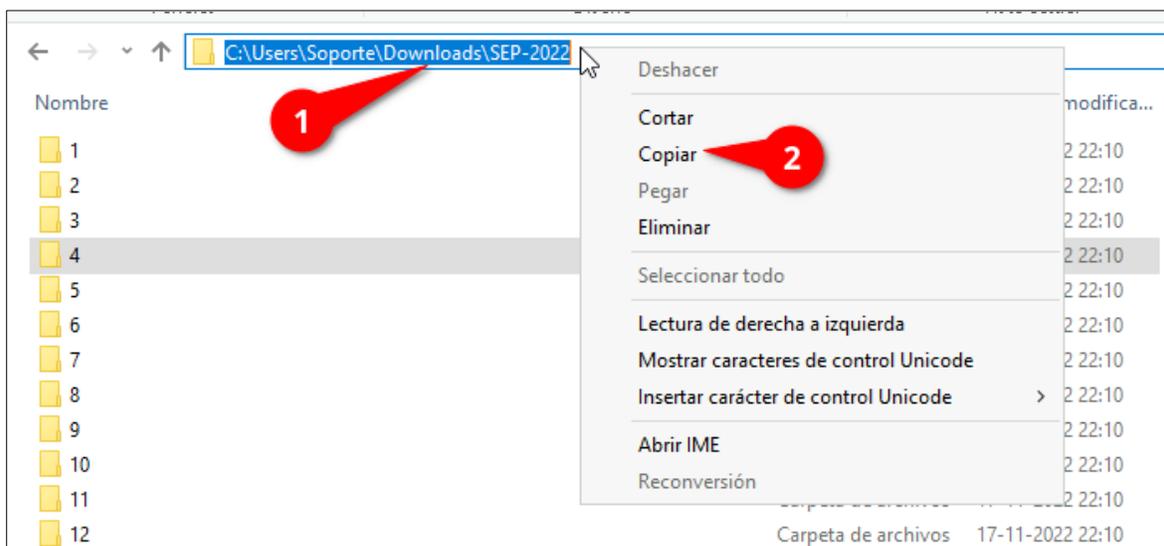
4. Luego de reiniciado Windows, ubicar la carpeta raíz en donde se encuentran los archivos PDF que se desea reducir, para este ejemplo `C:\Users\Soporte\Downloads\SEP-2022`.



UNIDAD DE RENDICIÓN ELECTRÓNICA DE CUENTAS



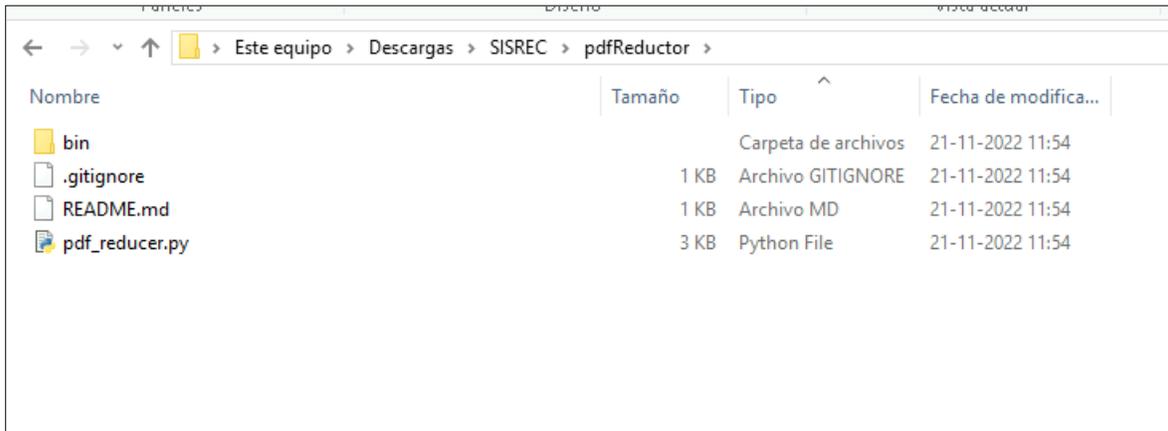
5. Hacer clic en la barra de direcciones (punto 1 de la imagen anterior) y copiar la ruta.



6. Ir a la carpeta en donde se encuentra el script pdfReductor, en nuestro caso `C:\Users\Soporte\Downloads\SISREC\pdfReductor`



UNIDAD DE RENDICIÓN ELECTRÓNICA DE CUENTAS



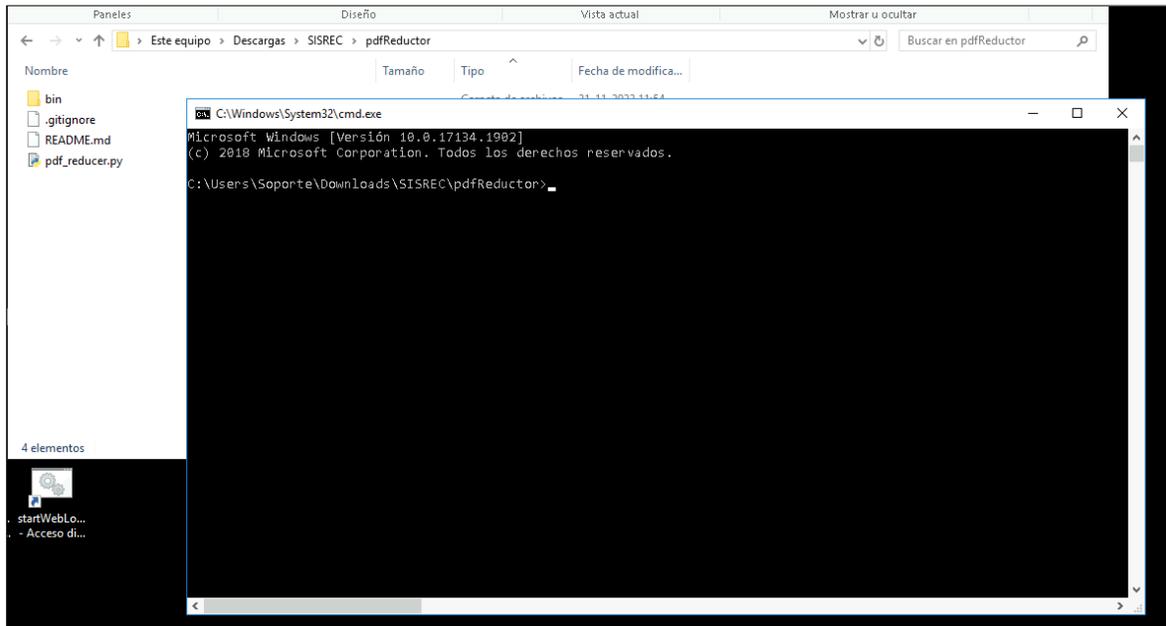
7. Luego, hacer clic en la barra de direcciones, borrar el contenido y escribir la palabra *cmd*



Lo que abrirá una ventana de *Símbolo del sistema*



UNIDAD DE RENDICIÓN ELECTRÓNICA DE CUENTAS

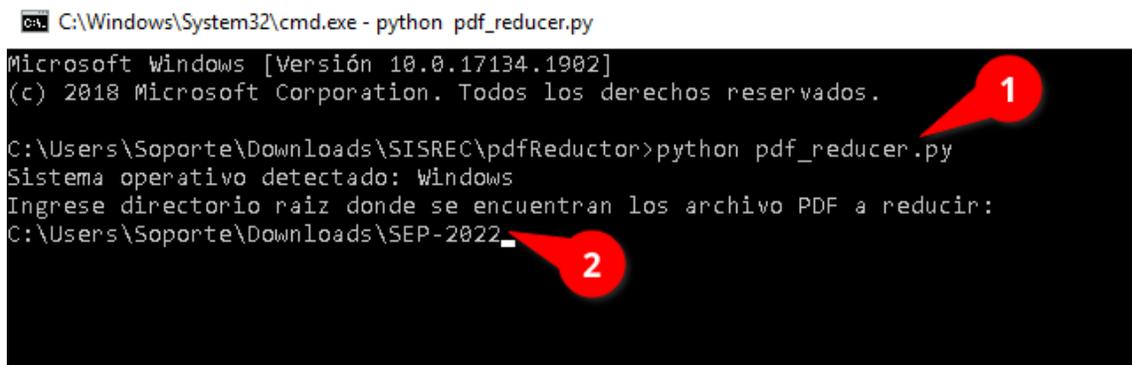


8. A continuación, en la ventana de *Símbolo del sistema*, escribir el comando (1):

```
python pdf_reduce.py
```

Y presionar *Enter*. Con esto, se le solicitará que ingrese la ruta de la carpeta raíz que contiene los archivos PDF y que fue copiada en el punto 5 de este manual (2), y presionar nuevamente la tecla *Enter*.

Puede volver al paso 5, y luego volver a esta ventana y presionar con el botón derecho del mouse sobre la ventana negra, y se pegará la ruta a la carpeta raíz; de lo contrario, deberá escribir la ruta manualmente.



9. Con esto, el script comenzará su ejecución y recorrerá cada una de las subcarpetas buscando archivos PDF y reduciéndolos.

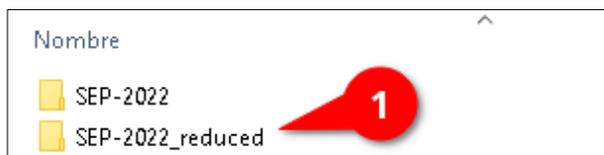


UNIDAD DE RENDICIÓN ELECTRÓNICA DE CUENTAS



```
CA:\Windows\System32\cmd.exe - python pdf_reducer.py
C:\Users\Soporte\Downloads\SISREC\pdfReductor>python pdf_reducer.py
Sistema operativo detectado: Windows
Ingrese directorio raíz donde se encuentran los archivo PDF a reducir:
C:\Users\Soporte\Downloads\SEP-2022
Reduciendo archivos...
Reduciendo archivo: C:\Users\Soporte\Downloads\SEP-2022\1\CE\Egreso 228122.pdf
1/5908 (0.02%)
Reduciendo archivo: C:\Users\Soporte\Downloads\SEP-2022\1\T\Contrato Daniela da Silva (indefinido).pdf
2/5908 (0.03%)
Reduciendo archivo: C:\Users\Soporte\Downloads\SEP-2022\1\T\Egreso 228122, Cartola.pdf
3/5908 (0.05%)
Reduciendo archivo: C:\Users\Soporte\Downloads\SEP-2022\1\T\Egreso 228122, Cheque N°109 Daniela Ribeiro da Silva.pdf
4/5908 (0.07%)
Reduciendo archivo: C:\Users\Soporte\Downloads\SEP-2022\1\T\Egreso 228122, Formulario Cheque N°109 Daniela Ribeiro da Si
5/5908 (0.08%)
Reduciendo archivo: C:\Users\Soporte\Downloads\SEP-2022\1\T\Egreso 228122, Liq. Agosto Daniela da Silva.pdf
6/5908 (0.1%)
Reduciendo archivo: C:\Users\Soporte\Downloads\SEP-2022\1\T\Impuesto F29 Agosto.pdf
7/5908 (0.12%)
Reduciendo archivo: C:\Users\Soporte\Downloads\SEP-2022\1\T\Libro de Remuneraciones Agosto 2022.pdf
8/5908 (0.14%)
```

10. Finalmente, terminado el proceso, junto a la carpeta raíz del punto 5, encontrará una carpeta con el mismo nombre con el posfijo *_reduced*, en nuestro ejemplo sería *C:\Users\Soporte\Downloads\SEP-2022_reduced*.



Esta nueva carpeta contiene la misma estructura que la carpeta original, y contiene los mismos archivos PDF pero en su versión reducida.

Consideraciones finales

La carpeta resultante (*_reduced*), es la que debe ser comprimida en ZIP para ser utilizada para el proceso de carga masiva que necesita ejecutar. Sin embargo, antes de comprimirla, debe cambiar su nombre para hacerla coincidir con el nombre del mes y el año necesarios, tal como lo solicita el proceso de carga masiva que se ejecutará en SISREC.