

Manual del Webservice “LeerArchivosCarpetaDrive”

1. Nombre del Webservice

LeerArchivosCarpetaDriveV3

2. Enlace al código original en Google Apps Script

[Ver código original en Google Apps Script](#)

3. Propietario del Webservice

homero.ortega@radiogis.uis.edu.co, cabcabrales1403@gmail.com y mariaalfonso8767@gmail.com

4. URL del Webservice

Bash

Copiar

https://script.google.com/macros/s/AKfycbxExzlV92_K04TR8kfzO1BbMvR-G3s7BqCTf6TVl6vQIDjIJQg1tmby9rrxr-UizQIX8Q/exec

5. Descripción del Webservice

El webservice LeerArchivosCarpetaDriveV3 es un servicio REST desarrollado en Google Apps Script que automatiza la lectura y extracción de texto de todos los archivos PDF contenidos en una carpeta específica de Google Drive.

A partir de esta versión (V3), el webservice NO mueve, elimina ni archiva los archivos procesados: todos los documentos permanecen en la carpeta original, permitiendo la gestión centralizada de los documentos y facilitando la manipulación y consulta continua por parte de los usuarios o sistemas externos.

Funcionamiento:

- **Selección de archivos:** El servicio recibe el parámetro `folder_id`, correspondiente a la carpeta de Google Drive a procesar. Obtiene todos los archivos con tipo MIME `application/pdf` ubicados en dicha carpeta.
- **Extracción de texto:** Cada PDF se convierte temporalmente a Google Docs para extraer el texto íntegro. El documento temporal se elimina inmediatamente después de la extracción, sin afectar el archivo original.
- **Gestión de archivos:** Los archivos PDF no se mueven, no se eliminan ni se renombran: permanecen inalterados en la carpeta de origen.
- **Respuesta:** El webservice devuelve un objeto JSON donde cada clave es el nombre del PDF y cada valor corresponde al texto extraído. Si ocurre un error, la respuesta es un mensaje de texto plano describiendo el problema.

Gracias a este web service, toda la gestión y manipulación de documentos PDF se realiza en una sola carpeta, lo que agiliza el acceso y optimiza el seguimiento de las todas las documentaciones.

6. Cambios en el funcionamiento (Historial de Cambios)

- **V3:** Ahora los archivos PDF procesados permanecen en la carpeta original/origen (“DOCUMENTOS”). No se mueven ni eliminan después de la extracción de texto, logrando tener un control total de los documentos en un solo sector.
- **V2:** Los archivos procesados se movían a otra carpeta (“YA_PROCESADOS”).
- **V1:** Los archivos procesados se eliminaban rotundamente de la carpeta (“DOCUMENTOS”).

7. Recursos para un ejemplo

- Carpeta en Google Drive:

URL de la carpeta de origen (DOCUMENTOS):

Copiar

https://drive.google.com/drive/folders/1bG5CINGwfEVc3DyFKlZ90_bV6xXH_0Tk

ID de la carpeta de origen (DOCUMENTOS):

Copiar

1bG5CINGwfEVc3DyFKlZ90_bV6xXH_0Tk

8. Sección de ejemplos.

a) Ejemplo para usar el Webservice en un Navegador

Para probar el servicio desde el navegador, ingresa la siguiente URL (asegúrate de que el parámetro folder_id sea el ID correcto):

Bash

Copiar

https://script.google.com/macros/s/AKfycbxExzIV92_K04TR8kfzO1BbMvR-G3s7BqCTf6TVl6vQIDjIJQg1tmby9rrxr-UizQIX8Q/exec

Al acceder, el webservice procesará los archivos PDF que se encuentran en la carpeta de origen llamada “DOCUMENTOS” y devolverá un JSON con sus nombres y contenidos extraídos.

b) Ejemplo Simple en Python

Python

Copiar

```
import requests

url =
"https://script.google.com/macros/s/AKfycbxExzlV92_K04TR8kfz
O1BbMvR-G3s7BqCTf6TVl6vQIDjIJQg1tmby9rrxr-UizQIX8Q/exec"
folder_id = "1bG5CINGwfEVc3DyFKlZ90_bV6xXH_0Tk"

response = requests.get(url, params={"folder_id": folder_id})
if response.status_code == 200:
    print("Respuesta del Webservice:")
    try:
        print(response.json())
    except ValueError:
        print("Respuesta no es JSON:", response.text)
else:
    print(f"Error {response.status_code}: {response.text}")
```

c) Ejemplo en Python: Función

```
WEB_SERVICE_URL="https://script.google.com/macros/s/AKfycbxExzIV  
92_K04TR8kfzO1BbMvR-G3s7BqCTf6TVl6vQIDjIJQg1tmby9rrxr-  
UizQIX8Q/exec"
```

```
FOLDER_ID = "1bG5CINGwfEVc3DyFKlZ90_bV6xXH_0Tk"
```

```
def load_docs_via_webservice(id_carpeta):
```

```
    """
```

```
    Llama al webservice 'LeerArchivosCarpetaDrive' y retorna un  
    diccionario  
    con los nombres de los archivos PDF y su contenido textual extraído.
```

```
    Parámetros:
```

```
        id_carpeta (str): ID de la carpeta en Google Drive.
```

```
    Retorna:
```

```
        dict: Diccionario con la estructura {nombre_archivo:  
        texto_extraído}.
```

```
    """
```

```
    params = {"folder_id": id_carpeta}
```

```
    response = requests.get(WEB_SERVICE_URL, params=params)
```

```
    if response.status_code == 200:
```

```
        try:
```

```
            return response.json() # Retorna el JSON parseado
```

```
        except ValueError:
```

```
            return {"error": "No se pudo parsear la respuesta como JSON",  
            "raw_response": response.text}
```

```
        else:
```

```
            raise Exception(f"Error al llamar al webservice:  
{response.status_code} - {response.text}")
```