📄 Instrucciones para Configurar OCR en Windows 11

# 1. Instalar Poppler

1. Ve al enlace: https://github.com/oschwartz10612/poppler-windows/releases  
2. Descarga el archivo .zip más reciente (por ejemplo: poppler-24.08.0-0.zip).  
3. Extrae su contenido en una ruta permanente como: C:\Apps\P\poppler-24.08.0

# 2. Instalar Tesseract

1. Ve al enlace: https://github.com/UB-Mannheim/tesseract/wiki  
2. Descarga e instala el archivo tesseract-ocr-w64-setup-5.x.x.exe.  
3. Durante la instalación, selecciona el idioma español (spa).  
4. Instala en: C:\Program Files\Tesseract-OCR

# 3. Ejecutar Script BAT para Variables de Entorno

1. Guarda el siguiente contenido como un archivo llamado configurar\_ocr\_entorno.bat  
2. Haz clic derecho sobre él y selecciona “Ejecutar como administrador”

@echo off  
setlocal  
  
:: Rutas necesarias  
set "TESSERACT\_PATH=C:\Program Files\Tesseract-OCR"  
set "POPLER\_PATH=C:\Apps\P\poppler-24.08.0\Library\bin"  
  
:: Mostrar título  
echo ======================================================  
echo Configurando variables de entorno para Tesseract y Poppler  
echo ======================================================  
echo.  
  
:: Agregar Tesseract si no está en PATH  
echo Verificando Tesseract...  
echo %PATH% | find /I "%TESSERACT\_PATH%" >nul  
if errorlevel 1 (  
 echo Agregando ruta de Tesseract al PATH...  
 setx PATH "%PATH%;%TESSERACT\_PATH%" /M  
) else (  
 echo Ruta de Tesseract ya está configurada.  
)  
  
:: Agregar Poppler si no está en PATH  
echo.  
echo Verificando Poppler...  
echo %PATH% | find /I "%POPLER\_PATH%" >nul  
if errorlevel 1 (  
 echo Agregando ruta de Poppler al PATH...  
 setx PATH "%PATH%;%POPLER\_PATH%" /M  
) else (  
 echo Ruta de Poppler ya está configurada.  
)  
  
echo.  
echo ✅ Configuración completada. Por favor:  
echo - Cierra esta ventana  
echo - Reinicia tu computadora o vuelve a iniciar sesión  
echo - Luego prueba en PowerShell: tesseract --version y pdftoppm -h  
pause  
endlocal

# 4. Instalar Librerías Python

Abre el Anaconda Prompt o una terminal de Spyder, y ejecuta:  
  
pip install pytesseract pdf2image python-docx PyPDF2 pandas  
  
O bien dentro de Spyder:  
  
!pip install pytesseract pdf2image python-docx PyPDF2 pandas

# 5. Verificación

En PowerShell, escribe:  
  
tesseract --version  
pdftoppm -h  
  
Y verifica que no haya errores.