Python Script to Bulk Rename TDS Certificate files

- Create a Folder called TDS at D:\
- Create a sub-folder called Certificates at D:\TDS\
- Copy all TDS Certificates downloaded from TRACES website to the Location "D:\TDS\Certificates"
- Download below file "TDS Certificates PDF RENAMER" and move to Location D:\TDS
- Run below commands from command line

```
pip install pymupdf
```

```
python d:\tds\tdsren.py
```

tdsren.py

```
# ============================================================
# https://www.planmytax.com/
#
# PlanMyTax TDS Certificate PDF Renamer
#
# Original:
#     ABC123.pdf
#
# Renamed To:
#     TDS Certificate - FY27 - Deductor - 2026.04 to 2026.06 - Deductee
- ABC123.pdf
#
# Requirements:
#     pip install pymupdf
# ============================================================

import fitz  # PyMuPDF
import re
import os
from pathlib import Path

# ============================================================
# CONFIGURATION
# ============================================================

PDF_FOLDER = r"D:\TDS\Certificates"

# ============================================================
# HELPERS
# ============================================================

def clean_filename(text):
    """Remove invalid Windows filename characters"""
```

```
    if not text:
        return ""

    text = re.sub(r'[<>:"/\\|?*]', '', text)
    text = re.sub(r'\s+', ' ', text)

    return text.strip()


def convert_date_to_period(date_str):
    """
    Converts:
        01-Apr-2026
    To:
        2026.04
    """

    months = {
        'JAN': '01',
        'FEB': '02',
        'MAR': '03',
        'APR': '04',
        'MAY': '05',
        'JUN': '06',
        'JUL': '07',
        'AUG': '08',
        'SEP': '09',
        'OCT': '10',
        'NOV': '11',
        'DEC': '12'
    }

    m = re.search(
        r'(\d{1,2})[-/ ]([A-Za-z]{3})[-/ ](\d{4})',
        date_str,
        re.I
    )

    if not m:
        return None

    year = m.group(3)
    month = months.get(m.group(2).upper())

    if not month:
        return None

    return f"{year}.{month}"
```

```
def derive_fy(period_from):
    """
    period_from = 2026.04
    returns FY27
    """

    if not period_from:
        return "FY??"

    year = int(period_from[:4])
    month = int(period_from[5:7])

    if month >= 4:
        return f"FY{str(year + 1)[-2:]}"
    else:
        return f"FY{str(year)[-2:]}"


def extract_text(pdf_file):

    text = ""

    pdf = fitz.open(pdf_file)

    for page in pdf:
        text += page.get_text("text")

    pdf.close()

    return text


# ============================================================
# FIELD EXTRACTION
# ============================================================

def extract_fields(text):

    deductor = ""
    deductee = ""
    period_from = ""
    period_to = ""

    # ------------------------------------------------------
    # DEDUCTOR
    # ------------------------------------------------------

    m = re.search(
        r'Name\s+and\s+Address\s+of\s+the\s+Deductor\s*(.*?)\n',
        text,
```

```
        re.I | re.S
    )

    if m:
        deductor = clean_filename(m.group(1))

    # ---------------------------------------------------------
    # DEDUCTEE
    # ---------------------------------------------------------

    m = re.search(
        r'Name\s+and\s+Address\s+of\s+the\s+Deductee\s*(.*?)\n',
        text,
        re.I | re.S
    )

    if m:
        deductee = clean_filename(m.group(1))

    # ---------------------------------------------------------
    # PERIOD
    # ---------------------------------------------------------

    m = re.search(
        r'Period\s+From\s+([^\n]+?)\s+To\s+([^\n]+)',
        text,
        re.I
    )

    if m:
        period_from = convert_date_to_period(m.group(1))
        period_to = convert_date_to_period(m.group(2))

    return deductor, deductee, period_from, period_to


# ============================================================
# MAIN
# ============================================================

folder = Path(PDF_FOLDER)

if not folder.exists():
    print(f"Folder not found: {PDF_FOLDER}")
    raise SystemExit

print(f"\nProcessing folder: {folder}\n")

for pdf_file in folder.glob("*.pdf"):
```

```
    try:

        # ------------------------------------------------
        # Skip already processed files
        # ------------------------------------------------

        if pdf_file.stem.startswith("TDS Certificate - "):
            print(f"SKIPPED (already renamed): {pdf_file.name}")
            continue

        text = extract_text(pdf_file)

        deductor, deductee, pfrom, pto = extract_fields(text)

        # ------------------------------------------------
        # Defaults
        # ------------------------------------------------

        if not deductor:
            deductor = "Unknown Client"

        if not deductee:
            deductee = "Unknown Supplier"

        # ------------------------------------------------
        # Financial Year
        # Derived from Period From
        # ------------------------------------------------

        fy = derive_fy(pfrom)

        # ------------------------------------------------
        # Period
        # ------------------------------------------------

        if pfrom and pto:

            if pfrom == pto:
                period = pfrom
            else:
                period = f"{pfrom} to {pto}"

        else:
            period = "Unknown Period"

        # ------------------------------------------------
        # Prevent excessively long filenames
        # ------------------------------------------------

        deductor = deductor[:50]
```

```
        deductee = deductee[:50]

        # ------------------------------------------------
        # Keep Original Filename
        # ------------------------------------------------

        original_name = pdf_file.stem

        prefix = (
            f"TDS Certificate - {fy}"
            f" - {deductor}"
            f" - {period}"
            f" - {deductee}"
        )

        new_name = (
            prefix
            + " - "
            + original_name
            + pdf_file.suffix
        )

        new_name = clean_filename(new_name)

        new_path = pdf_file.with_name(new_name)

        # ------------------------------------------------
        # Diagnostic Output
        # ------------------------------------------------

        print("\n-----------------------------------------")
        print("File      :", pdf_file.name)
        print("FY        :", fy)
        print("Deductor  :", deductor)
        print("Deductee  :", deductee)
        print("Period    :", period)
        print("-----------------------------------------")

        # ------------------------------------------------
        # Rename
        # ------------------------------------------------

        if new_path.exists():
            print("TARGET EXISTS - SKIPPED")
            continue

        os.rename(pdf_file, new_path)

        print("RENAMED TO:")
```

```
            print(new_name)

        except Exception as e:

            print(f"\nERROR: {pdf_file.name}")
            print(str(e))

print("\nCompleted.")
```

From:
https://planmytax.com/ - **PlanMyTax**

Permanent link:
https://planmytax.com/doku.php/python-script-to-bulk-rename-tds-certificate-files

Last update: **2026/06/22 08:03**