**## 🛠️ 1. Cài đặt thư viện**

Bạn sẽ cần cài đặt các thư viện tương ứng cho từng loại file. Mở Terminal hoặc Command Prompt và chạy lệnh duy nhất sau để cài đặt tất cả:

Bash

pip install pypdf python-docx openpyxl python-pptx

* **pypdf**: Dùng để đọc file PDF.
* **python-docx**: Dùng để đọc file Word (.docx).
* **openpyxl**: Dùng để đọc file Excel (.xlsx).
* **python-pptx**: Dùng để đọc file PowerPoint (.pptx).
* File **.txt** không cần thư viện ngoài, Python đã hỗ trợ sẵn.

**## 🐍 2. Mã nguồn hoàn chỉnh**

Đây là một script Python hoàn chỉnh. Bạn có thể lưu nó vào một file có tên file\_reader.py. Script này bao gồm các hàm riêng biệt để đọc từng loại file và một hàm chính để tự động chọn đúng hàm dựa trên đuôi file.

Python

import os

from docx import Document

from openpyxl import load\_workbook

from pptx import Presentation

from pypdf import PdfReader

# 📄 Hàm đọc file .txt

def read\_txt(file\_path):

"""Đọc nội dung từ file text (.txt)."""

try:

with open(file\_path, 'r', encoding='utf-8') as f:

return f.read()

except Exception as e:

return f"Lỗi khi đọc file TXT: {e}"

# 📝 Hàm đọc file .docx (Word)

def read\_docx(file\_path):

"""Đọc nội dung từ file Word (.docx)."""

try:

doc = Document(file\_path)

full\_text = [para.text for para in doc.paragraphs]

return '\n'.join(full\_text)

except Exception as e:

return f"Lỗi khi đọc file DOCX: {e}"

# 📊 Hàm đọc file .xlsx (Excel)

def read\_xlsx(file\_path):

"""Đọc nội dung từ tất cả các ô trong file Excel (.xlsx)."""

try:

workbook = load\_workbook(filename=file\_path)

full\_text = []

for sheet in workbook.worksheets:

full\_text.append(f"--- Sheet: {sheet.title} ---")

for row in sheet.iter\_rows():

row\_text = [str(cell.value) if cell.value is not None else "" for cell in row]

full\_text.append(" | ".join(row\_text))

return '\n'.join(full\_text)

except Exception as e:

return f"Lỗi khi đọc file XLSX: {e}"

# 🖥️ Hàm đọc file .pptx (PowerPoint)

def read\_pptx(file\_path):

"""Đọc nội dung text từ tất cả các slide trong file PowerPoint (.pptx)."""

try:

prs = Presentation(file\_path)

full\_text = []

for i, slide in enumerate(prs.slides):

slide\_text = []

for shape in slide.shapes:

if hasattr(shape, "text"):

slide\_text.append(shape.text)

if slide\_text: # Chỉ thêm slide có text

full\_text.append(f"--- Slide {i+1} ---")

full\_text.append('\n'.join(slide\_text))

return '\n\n'.join(full\_text)

except Exception as e:

return f"Lỗi khi đọc file PPTX: {e}"

# 📕 Hàm đọc file .pdf

def read\_pdf(file\_path):

"""Đọc nội dung text từ file PDF."""

try:

reader = PdfReader(file\_path)

full\_text = []

for page in reader.pages:

full\_text.append(page.extract\_text() or "")

return '\n'.join(full\_text)

except Exception as e:

return f"Lỗi khi đọc file PDF: {e}"

# 🤖 Hàm chính để đọc file bất kỳ

def read\_file(file\_path):

"""

Tự động xác định định dạng file và gọi hàm đọc tương ứng.

"""

if not os.path.exists(file\_path):

return f"Lỗi: File không tồn tại tại đường dẫn '{file\_path}'"

# Lấy đuôi file và chuyển thành chữ thường

file\_extension = os.path.splitext(file\_path)[1].lower()

if file\_extension == '.txt':

return read\_txt(file\_path)

elif file\_extension == '.docx':

return read\_docx(file\_path)

elif file\_extension == '.xlsx':

return read\_xlsx(file\_path)

elif file\_extension == '.pptx':

return read\_pptx(file\_path)

elif file\_extension == '.pdf':

return read\_pdf(file\_path)

else:

return f"Lỗi: Định dạng file '{file\_extension}' không được hỗ trợ."

# --- CÁCH SỬ DỤNG ---

if \_\_name\_\_ == "\_\_main\_\_":

# Thay thế 'path/to/your/file.ext' bằng đường dẫn thực tế đến file của bạn.

# Ví dụ: file\_path = "C:/Users/Admin/Documents/baocao.docx"

file\_path = "path/to/your/file.ext"

print(f"Đang đọc file: {file\_path}\n")

content = read\_file(file\_path)

print("--- NỘI DUNG FILE ---")

print(content)

print("---------------------")

**## 🚀 3. Cách sử dụng**

1. **Lưu file**: Lưu mã nguồn trên vào một file tên là file\_reader.py.
2. **Chuẩn bị file**: Đặt các file (PDF, DOCX, XLSX,...) mà bạn muốn đọc vào cùng thư mục với file file\_reader.py cho tiện, hoặc bạn có thể dùng đường dẫn tuyệt đối.
3. **Sửa đường dẫn**: Mở file file\_reader.py, tìm đến dòng file\_path = "path/to/your/file.ext" ở cuối cùng.
4. **Thay thế** "path/to/your/file.ext" bằng tên hoặc đường dẫn đầy đủ của file bạn muốn đọc.
   * Ví dụ 1 (file cùng thư mục): file\_path = "sample.docx"
   * Ví dụ 2 (đường dẫn tuyệt đối trên Windows): file\_path = "D:\\TaiLieu\\hopdong.pdf"
   * Ví dụ 3 (đường dẫn tuyệt đối trên macOS/Linux): file\_path = "/home/user/documents/data.xlsx"
5. **Chạy script**: Mở Terminal/Command Prompt, điều hướng đến thư mục chứa file và chạy lệnh:

Bash

python file\_reader.py

Nội dung của file sẽ được in ra trên màn hình terminal. Bạn có thể dễ dàng thay đổi đường dẫn để đọc các file khác nhau mà không cần sửa đổi gì thêm trong các hàm đọc.