

CHƯƠNG 7 FILES

PERSISTENCE

CHƯƠNG NÀY CHUYÊN TỪ LÀM VIỆC VỚI CPU VÀ BỘ NHỚ CHÍNH (MAIN MEMORY), NOI DỮ LIỆU MẤT KHI TẮT MÁY, SANG BỘ NHỚ PHỤ (SECONDARY MEMORY) NHƯ FILE VÀ USB, NOI DỮ LIỆU ĐƯỢC LUU TRỮ LÂU DÀI VÀ CÓ THỂ CHUYÊN GIỮA CÁC HỆ THỐNG. CHÚNG TA SẼ HỌC CÁCH ĐỌC VÀ GHI FILE VĂN BẢN TRƯỚC, SAU ĐÓ MỚI ĐẾN FILE CƠ SỞ DỮ LIỆU NHỊ PHÂN.

OPENING FILES

ĐỂ ĐỌC HOẶC GHI FILE, TA MỞ FILE BẰNG `open("FILENAME")`. HỆ ĐIỀU HÀNH SẼ KIỂM TRA FILE TỒN TẠI VÀ CẤP CHO TA MỘT FILE HANDLE, DÙNG ĐỂ THAO TÁC VỚI DỮ LIỆU. NẾU FILE KHÔNG TỒN TẠI, PYTHON SẼ BÁO LỖI `FileNotFoundException`. SAU NÀY CÓ THỂ DÙNG `TRY/EXCEPT` ĐỂ XỬ LÝ LỖI NÀY.

```
# Mở file
fname = input("Enter the file name: ")
try:
    fhand = open(fname, "r") # mở file ở chế độ đọc
except FileNotFoundError:
    print("File cannot be opened:", fname)
    exit()

# Đọc và in nội dung từng dòng
for line in fhand:
    print(line.strip()) # loại bỏ ký tự xuống dòng

fhand.close() # đóng file
```

TEXT FILES AND LINES

FILE VĂN BẢN LÀ CHUỖI CÁC DÒNG, MỖI DÒNG KẾT THÚC BẰNG KÝ TỰ NEWLINE `\n`. TRONG PYTHON, `\n` LÀ MỘT KÝ TỰ DUY NHẤT GIÚP PHÂN TÁCH DÒNG; KHI IN RA, NÓ XUỐNG DÒNG. VÍ DỤ: "HELLO\nWORLD!" SẼ IN THÀNH HAI DÒNG.

```
# Ví dụ chuỗi với ký tự xuống dòng
stuff = "Hello\nWorld!"
print(stuff)
# Output:
# Hello
# World!

# Đếm số ký tự trong chuỗi có ký tự newline
stuff2 = "X\nY"
print(len(stuff2)) # Output: 3
```

READING FILES

DÙNG FILE HANDLE TRONG PYTHON, TA CÓ THỂ ĐỌC FILE TỪNG DÒNG BẰNG FOR LOOP ĐỂ ĐÉM DÒNG MÀ KHÔNG TỒN NHIỀU BỘ NHỚ, VÌ MỖI DÒNG ĐƯỢC ĐỌC, XỬ LÝ, RỒI LOẠI BỎ. NẾU FILE NHỎ, CÓ THỂ DÙNG READ() ĐỂ ĐỌC TOÀN BỘ NỘI DUNG VÀO MỘT CHUỖI.

```
# Đếm số dòng trong file
fhand = open('mbox-short.txt')
count = 0
for line in fhand:
    count += 1
print("Line Count:", count)

# Đọc toàn bộ nội dung file nếu file nhỏ
fhand = open('mbox-short.txt')
inp = fhand.read()
print("Số ký tự trong file:", len(inp))
```

LETTING THE USER CHOOSE THE FILE NAME

THAY VÌ SỬA CODE MỖI LẦN, TA CÓ THỂ HỎI NGƯỜI DÙNG NHẬP TÊN FILE, SAU ĐÓ MỞ FILE ĐÓ VÀ ĐÉM CÁC DÒNG BẮT ĐẦU BẰNG "SUBJECT:". ĐIỀU NÀY GIÚP CHƯƠNG TRÌNH DÙNG ĐƯỢC CHO NHIỀU FILE KHÁC NHAU MÀ KHÔNG CẦN CHỈNH SỬA CODE.

```
fname = input("Enter the file name: ")
fhand = open(fname)
count = 0
for line in fhand:
    if line.startswith("Subject:"):
        count += 1
print("There were", count, "subject lines in", fname)
```

GLOSSARY

CATCH: NGĂN LỖI LÀM CHƯƠNG TRÌNH DỪNG BẰNG TRY VÀ EXCEPT.

NEWLINE: KÝ TỰ ĐẶC BIỆT ĐÁNH DẤU KẾT THÚC DÒNG.

PYTHONIC: CÁCH LÀM GỌN, HIỆU QUẢ TRONG PYTHON.

QUALITY ASSURANCE (QA): NHÓM ĐẢM BẢO CHẤT LƯỢNG PHẦN MỀM.

TEXT FILE: CHUỖI KÝ TỰ LUU TRÊN BỘ NHỚ VĨNH VIỄN.