

Chương 7.

Reading Files

- Trước khi đọc nội dung, cần phải cho Python biết file nào sẽ làm việc và thao tác gì với file đó.
- Sử dụng hàm **open()** để thực hiện việc này.
- Hàm open() sẽ trả về một "**file handle**" (tay cầm file) - là một biến dùng để thực hiện các thao tác trên file, tương tự như việc "File -> Open" trong trình xử lý văn bản.
- Cú pháp:** handle = open(filename, mode). filename là chuỗi ký tự (string) tên file.mode là tùy chọn, nên là 'r' nếu đọc file (read) và 'w' nếu ghi file (write).
- Nếu file không tồn tại, sẽ xảy ra lỗi.
- FileNotFoundException.

Opening a File

- File Handle như một Chuỗi:** Một file handle mở để đọc có thể được coi là một chuỗi các chuỗi ký tự, trong đó mỗi dòng trong file là một chuỗi trong chuỗi đó.
Đọc File Từng Dòng: Có thể dùng vòng lặp **for** để lặp qua từng dòng của file handle.
- Đếm Dòng:** Có thể mở file, dùng vòng lặp for để đọc từng dòng và tăng biến đếm (count) lên 1.
- Đọc Toàn Bộ File:** Có thể đọc toàn bộ file (bao gồm cả ký tự \n) vào một chuỗi đơn bằng phương thức **.read()**.

File Processing

- Một file văn bản có thể được xem là một chuỗi các dòng (sequence of lines).
- Ký tự đặc biệt "**newline**" (\n) được sử dụng để báo hiệu kết thúc một dòng.
- File văn bản có ký tự \n ở cuối mỗi dòng.

- Tìm kiếm cơ bản:** Sử dụng câu lệnh if trong vòng lặp for để chỉ in ra các dòng đáp ứng tiêu chí nhất định. (ví dụ: if line.startswith('From:'):).
- Lỗi dòng trống (\n):** Khi in các dòng từ file, thường xuất hiện thêm dòng trống do mỗi dòng đã có sẵn \n và lệnh print() cũng thêm một \n.
- Khắc phục lỗi dòng trống:** Dùng phương thức **.rstrip()** từ thư viện string để loại bỏ khoảng trắng (bao gồm cả \n) ở bên phải chuỗi.
- Bỏ qua dòng với continue:** Có thể sử dụng câu lệnh **continue** để bỏ qua một dòng không đáp ứng tiêu chí và chuyển sang dòng tiếp theo.
- Sử dụng in để chọn dòng:** Dùng toán tử **in** để tìm kiếm một chuỗi con bất kỳ trong một dòng để làm tiêu chí chọn.

- Có thể nhắc người dùng nhập tên file bằng hàm **input()**.
- Sử dụng khối **try và except** để xử lý trường hợp người dùng nhập tên file không hợp lệ hoặc file không thể mở được, tránh cho chương trình bị sập.