

TÓM LƯỢC BÀI GIẢNG NHẬP MÔN LẬP TRÌNH

(Vũ Quốc Hoàng, vqhoang@fit.hcmus.edu.vn, FIT-HCMUS, 2020)

BÀI 9A THAO TÁC VỚI TẬP TIN

Chủ đề

- Tập tin văn bản

Tài liệu

[1] Vũ Quốc Hoàng, *Bí kíp luyện Lập trình C (Quyển 1)*, hBook, 2017.

[2] Vũ Quốc Hoàng, *Bí kíp luyện Lập trình nhập môn với Python*, hBook, 2020.

Đọc tài liệu

- Đọc kĩ: Bài 4.7 [1]
- Đọc (và thực hành) thêm: Bài 17 [2]

Kiến thức

- Dữ liệu có thể được để: trên các **thanh ghi** và **vùng nhớ cache** của bộ vi xử lý, trong **bộ nhớ chính** của chương trình (main/primary memory, RAM) và ở **bộ nhớ ngoài** (secondary storage, auxiliary memory) như đĩa cứng, đĩa CD, DVD, đĩa flash (USB), ... Dữ liệu trên bộ nhớ ngoài thường được tổ chức thành các **tập tin** (file).
- Dữ liệu trên tập tin (và bộ nhớ ngoài) có ưu điểm là: thường trú (nghĩa là không mất đi khi đóng chương trình hay tắt máy), dung lượng lớn và có thể được chia sẻ, sao chép giữa các lần chạy chương trình, giữa các chương trình, các máy, người dùng, ...
- Dạng tập tin đơn giản nhất là **tập tin văn bản** (text file) với dữ liệu là dãy các kí tự, tức là chuỗi.
- **Đường dẫn tập tin** (file path), bao gồm **tên tập tin** (file name) với **phần mở rộng** thường là .txt, là chuỗi giúp định vị tập tin trên bộ nhớ ngoài.
- Chương trình chỉ có thể làm việc (trực tiếp) với dữ liệu trên bộ nhớ chính. Do đó ta cần đưa dữ liệu vào/ra giữa bộ nhớ chính và bộ nhớ ngoài khi làm việc với tập tin. Mô hình chung là: chương trình **đọc** (read) dữ liệu từ tập tin vào bộ nhớ chính, xử lý dữ liệu trên bộ nhớ chính, rồi **ghi** (write) dữ liệu trên bộ nhớ chính ra tập tin.
- **Con trỏ file** (file pointer) giúp quản lý tập tin trong chương trình. Chương trình cần **mở** (open) tập tin với **chế độ mở** phù hợp (đọc, ghi, ...), thao tác tương ứng (đọc, ghi, ...) và **đóng** (close) tập tin.
- Ta có thể làm việc với nhiều dữ liệu (chẳng hạn nhiều số) trong tập tin bằng kỹ thuật dùng số lượng (tương tự việc dùng thông tin chiều dài đi kèm với mảng) hoặc kỹ thuật nhận diện việc đọc hết nội dung tập tin (tương tự việc dùng kí tự NULL để đánh dấu kết thúc chuỗi). (Các kĩ thuật khác sẽ được học sau.)
- **Điêu hướng** (redirect) là kĩ thuật chuyển việc xuất thay vì ra màn hình (stdout thông thường) sẽ ra tập tin hoặc chuyển việc nhập thay vì từ bàn phím (stdin thông thường) sẽ từ tập tin.

Kĩ năng

- Biết cách tạo, xóa, mở, đọc, gõ, lưu nội dung tập tin văn bản như một người dùng bình thường (chẳng hạn biết dùng Notepad, Notepad++, ...)
- Thuần thục các bước làm việc cơ bản với tập tin: mở, thao tác (đọc, ghi, ...), đóng
- Biết cách tra cứu và vận dụng các hàm trong thư viện chuẩn C cho các thao tác trên tập tin văn bản (stdio.h)
- Biết cách viết chương trình thao tác đơn giản (đọc, ghi) trên tập tin văn bản
- Biết cách xử lý tập tin chứa nhiều dữ liệu bằng kỹ thuật dùng số lượng hay kỹ thuật đọc hết nội dung tập tin
- Vận dụng kỹ thuật điều hướng cho một số tình huống hay gặp
- Vận dụng tập tin (văn bản) để lưu trữ, tổ chức, xử lý dữ liệu cho các tình huống hay gặp

Lưu ý

- Kỹ thuật lưu trữ và xử lý dữ liệu trên bộ nhớ ngoài (tập tin) là rất quan trọng, hơn nữa bài này đặt nền móng cho các thao tác phức tạp hơn trên tập tin trong các bài sau, nên sinh viên phải nắm vững và thành thạo

Bài tập

1. Tìm một nguồn (web) tin cậy và tra cứu thư viện chuẩn C cho các kiểu dữ liệu, biến, hằng, hàm thao tác trên tập tin (stdio.h).
2. Các bài tập 4.7 [1].
3. Làm thêm các bài Bài 17 [2].