



# **CẤU TRÚC DỮ LIỆU VÀ GIẢI THUẬT**

## **CHƯƠNG 1: TỔNG QUAN VỀ GIẢI THUẬT VÀ CẤU TRÚC DỮ LIỆU**

# Nội dung

- 1.1. Vai trò của Cấu trúc dữ liệu trong một đề án tin học
- 1.2. Các tiêu chuẩn đánh giá dữ liệu
- 1.3. Kiểu dữ liệu
- 1.4. Đánh giá độ phức tạp của giải thuật

# VAI TRÒ CỦA CẤU TRÚC DỮ LIỆU

- Dữ liệu:
  - Không phần mềm nào là không có dữ liệu!
  - Việc chọn dữ liệu liên quan đến chất lượng chương trình (tốc độ xử lý, dung lượng, số dòng lệnh...)
- Thuật toán – Giải thuật – Thuật giải
  - **Là tập hợp (dãy) hữu hạn các chỉ thị (hành động) được định nghĩa rõ ràng nhằm giải quyết một bài toán cụ thể nào đó.**
- Minh họa bằng ngôn ngữ tự nhiên (natural language), bằng sơ đồ (flow chart) hoặc bằng mã giả (pseudo code)
- Mã giả bằng tựa C hoặc Pascal thường được sử dụng

# VAI TRÒ CỦA CẤU TRÚC DỮ LIỆU

■ Quan hệ giữa CTDL và GT:

- Cấu trúc dữ liệu + Giải thuật (+Giao diện) = Chương trình
- Không thể thiếu 1 trong hai đối tượng
- Việc tạo chương trình chỉ là vấn đề lựa chọn ngôn ngữ

# ĐÁNH GIÁ CTDL & GT

- Tiêu chuẩn đánh giá CTDL:
  - tiết kiệm tài nguyên (bộ nhớ trong),
  - phản ánh đúng thực tế của bài toán,
  - dễ dàng trong việc thao tác dữ liệu.
- Đánh giá độ phức tạp thuật toán:
  - Ước lượng thời gian thực hiện  $T(n)$  để so sánh tương đối
  - Xét hai thời gian  $T_{min}$  thấp nhất và  $T_{max}$  cao nhất để tính  $T_{avg}$

# Kiểu dữ liệu

## ■ Kiểu dữ liệu T có hai thành phần:

- Tập giá trị V
- Tập phép toán O
- $T = \langle V, O \rangle$

Mỗi kiểu phải có tên (định danh – Identifier) và độ lớn mỗi phần tử lưu trữ trong bộ nhớ máy tính, tính bằng B(Byte)

## ■ Các kiểu dữ liệu cơ sở

- | □ Kiểu số nguyên   | không dấu        | có dấu                  |
|--|------------------|-------------------------|
| ■ 1B   | 0->255           | -128->127               |
| ■ 2B   | 0->65535         | -32768->32767           |
| ■ 4B   | 0-> $2^{32} - 1$ | $-2^{31} -> 2^{31} - 1$ |
| ■ Phép toán: $O = \{+, -, *, /, \text{DIV}, \text{MOD}, \%, \text{so sánh}, \dots\}$ |                  |                         |

# Kiểu dữ liệu

- Kiểu số thực
  - 4B, 6B, 8B, 10B
  - Phép toán:  $O = \{+, -, *, /, \%, \text{so sánh}, \dots\}$
- Kiểu ký tự:
  - 1B, 2B (Unicode)
  - Phép toán:  $O = \{+, -, \text{so sánh}, \dots\}$
- Kiểu chuỗi (xâu) ký tự:
  - Độ lớn phụ thuộc ngôn ngữ
  - Phép toán:  $O = \{+, \&, \text{so sánh}, \dots\}$
- Kiểu logic (luận lý)
  - Độ lớn 1B có hai giá trị True, False
  - Phép toán:  $O = \{\text{AND}, \text{OR}, \text{NOT}, \text{XOR}, \text{so sánh}\}$

# Kiểu dữ liệu

- Các kiểu dữ liệu có cấu trúc
  - Xây dựng dựa trên các kiểu đã có
  - Có hai loại thông thường: Mảng và bản ghi (cấu trúc)
- Dữ liệu kiểu con trỏ:
  - Quản lý địa chỉ bộ nhớ
  - Có hai loại: Near (2B) và Far (4B)



# ĐÁNH GIÁ ĐỘ PHỨC TẠP CỦA THUẬT TOÁN

## ■ Các bước đánh giá

- Xem xét kích thước dữ liệu vào (vd:  $n = ?$ ), thời gian chạy là một hàm của  $n$ , ký hiệu là  $T(n)$
- Tách biệt các thao tác trừu tượng của thuật toán để xác định các phép toán phụ thuộc tốc độ xử lý. Không thể dựa vào các phép toán này để đánh giá thuật toán.
- Tìm ra các giá trị  $T(n)$  trong t/h xấu nhất, tốt nhất và trung bình.
- Ký hiệu  $O(f(n))$  là để biểu diễn độ phức tạp của thuật toán.

## ■ Sự phân lớp các thuật toán:

- $O(1)$ : Thời gian chạy là hằng số (không phụ thuộc  $n$ )
- $O(\log(n))$ : Thời gian là hàm Logarit của  $n$
- $O(n)$ : Thời gian tuyến tính

# ĐÁNH GIÁ ĐỘ PHỨC TẠP CỦA THUẬT TOÁN

- Sự phân lớp các thuật toán:
  - $O(n\log(n))$ ,  $O(n^2)$ ,  $O(n^3)$ : Thời gian đa thức
  - $O(2^n)$ ,  $O(n!)$ ...: Thời gian hàm mũ
- Phân tích trường hợp trung bình
  - T/h trung bình là t/h dữ liệu được cho ngẫu nhiên
  - Sử dụng các nguyên lý cộng và nhân để xác định số lần thực hiện các phép toán
  - Xác suất xuất hiện các đối tượng được coi là như nhau trong các phép toán

## BÀI TẬP