

Cấu trúc dữ liệu & Giải thuật (Data Structures and Algorithms)

Các khái niệm cơ bản



Nguyễn Tri Tuấn
Khoa CNTT – ĐH.KHTN.Tp.HCM
Email: nttuan@fit.hcmus.edu.vn



Nội dung

- 1 Kiểu dữ liệu (Data Type).....•
- 2 Kiểu dữ liệu cơ bản (Basic Data Type).....•
- 3 Kiểu dữ liệu có cấu trúc (Structured Data Type).....•
- 4 Kiểu dữ liệu trừu tượng (ADT – Abstract Data Type).....•
- 5 Cấu trúc dữ liệu (Data structure).....•
- 6 Đánh giá Cấu trúc dữ liệu.....•



Kiểu dữ liệu (1)

- Hãy viết ra ít nhất 5 kiểu dữ liệu mà bạn biết.
 - Mô tả ngắn gọn các đặc điểm của mỗi kiểu dữ liệu



Kiểu dữ liệu (2)

■ Ví dụ:

- Kiểu số nguyên (int)
- Kiểu ký tự (char)
- Kiểu chuỗi (string)
- Kiểu mảng (array)
- ...

■ Định nghĩa tổng quát “Kiểu dữ liệu”

$$\mathbf{T} = \langle \mathbf{V}, \mathbf{O} \rangle$$

- V (Values - miền giá trị): tập hợp các *giá trị* mà kiểu T có thể nhận
- O (Operators – các thao tác): tập hợp các *thao tác* cơ bản được định nghĩa *trên V*



Kiểu dữ liệu (3)

■ Ví dụ

- T = short int (2 bytes)
 - $V = \{-32,768 \dots +32,767\}$
 - $O = \{+, -, *, \text{div}, \text{mod}, >, >=, <, <=, ==, !=, <<, >>\}$
- T = int (4 bytes)
 - $V = \{-2,147,483,648 \dots 2,147,483,647\}$
 - $O = \{+, -, *, \text{div}, \text{mod}, >, >=, <, <=, ==, !=, <<, >>\}$
- T = unsigned char (1 bytes)
 - $V = \{0 \dots 255\}$
 - $O = \{+, -, *, \text{div}, \text{mod}, >, >=, <, <=, ==, !=, <<, >>\}$



Kiểu dữ liệu cơ bản (1)

- Các ngôn ngữ lập trình (C/C++/Java,...) đều cung cấp sẵn các kiểu dữ liệu cơ bản để người lập trình sử dụng
 - Các kiểu số nguyên: short int, int, long, char
 - Kiểu logic: bool
 - Các kiểu số thực: float, double



Kiểu dữ liệu cơ bản (2)

| Kiểu dữ liệu | Kích thước (size) | Miền giá trị |
|-------------------------------|-------------------|--------------|
| bool | 1 byte | ? |
| char, unsigned char | 1 byte | ? |
| short, unsigned short | 2 bytes | ? |
| int, unsigned int | 4 bytes | ? |
| long, unsigned long | 4 bytes | ? |
| long long, unsigned long long | 8 bytes | ? |
| float | 4 bytes | ? |
| double | 8 bytes | ? |



Kiểu dữ liệu có cấu trúc (1)

- Người lập trình cũng có thể xây dựng các kiểu dữ liệu mới bằng cách kết hợp các kiểu cơ bản thành một kiểu cấu trúc:
 - Kiểu mảng: array
 - Kiểu chuỗi ký tự: string
 - Kiểu struct
 - Kiểu tập hợp: enum
 - Kiểu union



Kiểu dữ liệu có cấu trúc (2)

- Kiểu array:

- VD. `int NumList[100];` // array gồm 100 int. Size = ?

- Kiểu string:

- VD. `char Name[30];` // array gồm 30 char. Size = ?

- Kiểu struct:

- VD.

```
struct DATE {  
    unsigned short int Year, Month, Day;  
}; // Size = ?  
struct PERSON {  
    char CardID[9];           // số CMND  
    char Name[30];  
    struct DATE Birthday;  
    float Weight;  
}; // Size = ?
```



Kiểu dữ liệu có cấu trúc (3)

■ Kiểu enum:

```
enum BOOLEAN
{
    false,    // false = 0, true = 1
    true
};
enum BOOLEAN isCorrect = true; // giá trị của biến = 1

enum WEEKDAYS                // tập hợp các ngày trong tuần
{
    sunday,                // sunday=0, monday=1, tuesday=2, ...
    monday,
    tuesday,
    wednesday,
    thursday,
    friday,
    saturday
};
enum WEEKDAYS today = thursday;
```



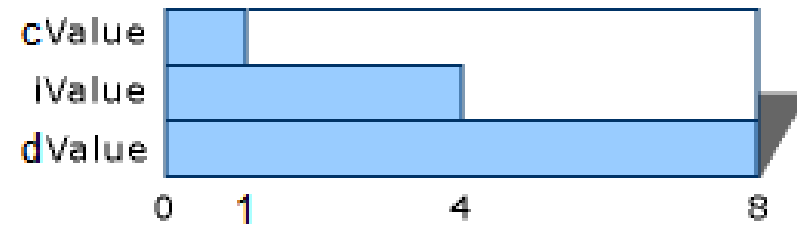
Kiểu dữ liệu có cấu trúc (4)

■ Kiểu union:

```
// using_a_union.cpp  
#include <stdio.h>
```

```
union NumericType  
{  
    char    cValue;  
    int     iValue;  
    double  dValue;  
}; // Size = 8 bytes
```

```
int main()  
{  
    union NumericType Values;  
    Values.iValue = 1000;  
    printf("%d\n", Values.iValue);  
    Values.dValue = 3.1416;  
    printf("%f\n", Values.dValue);  
}
```





Nội dung

- 1 Kiểu dữ liệu (Data Type).....●
- 2 Kiểu dữ liệu cơ bản (Basic Data Type).....●
- 3 Kiểu dữ liệu có cấu trúc (Structured Data Type).....●
- 4 Kiểu dữ liệu trừu tượng (ADT – Abstract Data Type).....●
- 5 Cấu trúc dữ liệu (Data structure).....●
- 6 Đánh giá Cấu trúc dữ liệu.....●



Kiểu dữ liệu trừu tượng (1)

■ Định nghĩa ADT

- Là một tập các giá trị, cùng với các thao tác liên quan
- Không chỉ rõ cách thức cài đặt cụ thể (độc lập với cách thức cài đặt)

■ Ví dụ:

- Stack ADT
 - Tập các phần tử
 - Các thao tác: push, pop, peak
- Có nhiều cách cài đặt Stack ADT:
 - Cài đặt dùng mảng 1 chiều
 - Cài đặt dùng danh sách liên kết



Kiểu dữ liệu trừu tượng (2)

- Hãy cho 3 ví dụ về ADT mà bạn biết
 - Mô tả các thao tác cơ bản
 - Nêu ít nhất 2 cách cài đặt cho mỗi ADT



Cấu trúc dữ liệu (1)

- Là cách thức *tổ chức* (organizing) và *lưu trữ* (storing) dữ liệu trong bộ nhớ (memory) để mang lại hiệu quả khi thi hành thuật toán
- Cấu trúc dữ liệu là cách thức cài đặt của ADT
 - Danh sách liên kết (Linked list), hàng đợi (Queue), ngăn xếp (Stack), cây (Tree), từ điển (Dictionary), Heap,...
- *External memory data structure*



Cấu trúc dữ liệu (2)

- Mỗi cấu trúc dữ liệu sẽ thích hợp cho một ứng dụng cụ thể
 - B-cây thích hợp để dùng cho database
 - Trình biên dịch thường dùng bảng băm (Hash table) để tìm kiếm
 - Bảng băm cũng thường dùng cho ứng dụng Từ điển (dictionary)
 - Hàng đợi (Queue) dùng cho ứng dụng phân phối hàng hoá
 - ...



Nội dung

- 1 Kiểu dữ liệu (Data Type).....•
- 2 Kiểu dữ liệu cơ bản (Basic Data Type).....•
- 3 Kiểu dữ liệu có cấu trúc (Structured Data Type).....•
- 4 Kiểu dữ liệu trừu tượng (ADT – Abstract Data Type).....•
- 5 Cấu trúc dữ liệu (Data structure).....•
- 6 Đánh giá Cấu trúc dữ liệu.....•



Đánh giá Cấu trúc dữ liệu (1)

- Một cấu trúc dữ liệu được gọi là thích hợp cho một ứng dụng (A) nếu thoả được các điều kiện sau:
 - Lưu trữ đầy đủ và đúng đắn dữ liệu của A
 - Dễ dàng truy xuất và xử lý
 - Tiết kiệm bộ nhớ



Đánh giá Cấu trúc dữ liệu (2)

- Tính đầy đủ và đúng đắn:
 - VD1. dữ liệu cần lưu là “điểm trung bình”
int DiemTB;
char DiemTB;
float DiemTB;
 - VD2. dữ liệu cần lưu là “ngày” [1-31]
int Ngay;
short int Ngay;
unsigned short int Ngay;
float Ngay;
 - VD3. dữ liệu cần lưu là “năm”
unsigned char Nam;
unsigned int Nam;
unsigned short int Nam;



Đánh giá Cấu trúc dữ liệu (3)

- Tính đầy đủ và đúng đắn:
 - VD4. dữ liệu cần lưu là “đơn giá mặt hàng (VND)”
unsigned short int Dongia;
unsigned int Dongia;
float Dongia;
unsigned long long Dongia;
 - VD5. dữ liệu cần lưu là “đơn giá mặt hàng (USD)”
unsigned short int Dongia;
unsigned int Dongia;
float Dongia;



Đánh giá Cấu trúc dữ liệu (4)

- Tính dễ dàng truy xuất và xử lý
 - VD. dữ liệu cần lưu là “ngày sinh”
char Ngaysinh[8]; // ddmmyyyy
char Ngaysinh[8]; // yyymdd
struct DATE Ngaysinh;
- Tính tiết kiệm bộ nhớ
 - Xem VD. trên



Q & A

