

# Bài 5.1: Giới thiệu về cây

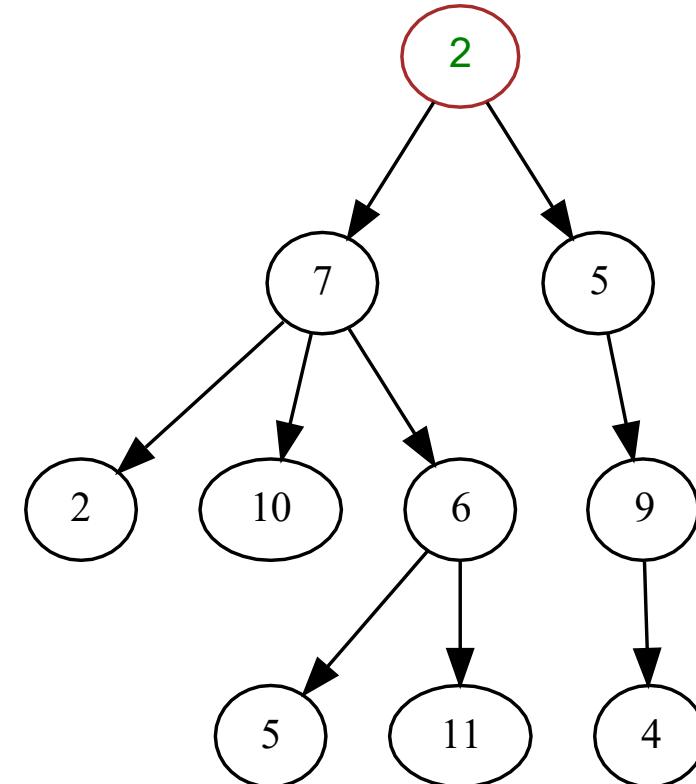
---

- ✓ Định nghĩa
- ✓ Ứng dụng của cây
- ✓ Các thuật ngữ
- ✓ Các thao tác với cây
- ✓ Biểu diễn của cây

# Định nghĩa

- Một cấu trúc dữ liệu cây(tree) có thể được định nghĩa một cách đệ quy bởi tập hợp của các node bắt đầu từ node gốc, trong đó mỗi node là một cấu trúc dữ liệu bao gồm một giá trị và một danh sách các tham chiếu tới các node con với ràng buộc rằng không có tham chiếu nào bị trùng lặp và không có tham chiếu nào trở ngược đến node gốc.
- Trong khoa học máy tính, cây là một kiểu dữ liệu trừu tượng được sử dụng rộng rãi mô phỏng cấu trúc dữ liệu phân cấp, với giá trị gốc và các cây con cùng với node cha của nó, biểu diễn dưới dạng tập hợp các nút được liên kết với nhau.
- Cây thường được biểu diễn theo thứ tự từ trên xuống: node gốc sau đó đến các node con, sau cùng là node lá.

# Ví dụ về cây



# Ứng dụng

- Biểu diễn các dữ liệu phân cấp như:
  - Tài liệu JSON, YAML
  - Cây cú pháp trừu tượng cho ngôn ngữ máy tính
  - Cây phân tích cú pháp cho ngôn ngữ con người
  - Mô hình hóa đối tượng tài liệu XML, HTML
- Cây tìm kiếm lưu trữ dữ liệu theo cách làm cho thuật toán tìm kiếm hiệu quả có thể thực hiện được thông qua việc duyệt cây.
- Biểu diễn danh sách dữ liệu đã được sắp xếp.
- Cây lưu trữ Barnes-Hut dùng để mô phỏng các thiên hà.

# Các thuật ngữ

- Một node: là một cấu trúc có thể chứa một giá trị hay điều kiện hoặc biểu diễn một cấu trúc dữ liệu tách biệt.
- Node con: là các node phía dưới node khác trong cây. Một node có thể có 0, 1 hoặc nhiều node con.
- Node cha: là node liền kề phía trên có liên kết trực tiếp đến node phía dưới(con). Một node có tối đa 1 cha và có thể có nhiều node tổ tiên.
- Node tổ tiên: các node cha của node cha.
- Node anh em: các node có cùng node cha.
- Node cành hay còn gọi node trong: là bất kì node nào có node con.
- Node lá hay còn gọi node ngoài: là node không có node con.
- Node gốc(root): là node trên cùng của cây. Node gốc không có node cha. Để đến được các node con phải biết node gốc.

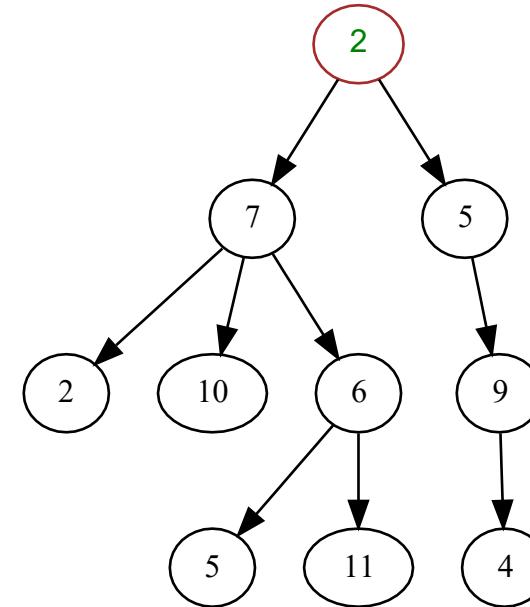
# Các thuật ngữ

- Cạnh: đường nối hay liên kết giữa node cha và node con.
- Chiều cao của 1 node: là độ dài lớn nhất từ node đó xuống tới node lá của nó. Chiều cao của node gốc là chiều cao của cây. Node lá có chiều cao bằng 0.
- Độ sâu của một node: là độ dài từ node gốc xuống tới node đó. Node gốc có chiều sâu bằng 0.
- Cây rỗng: cây không có node nào, cây này có chiều cao là -1.
- Cây con: là cây chứa các node cấu thành của cây T. Các node của cây con là hậu duệ trong T.
- Hàng xóm: node cha hoặc node con.
- Hậu duệ: một node có thể đi tới được bằng cách duyệt từ node cha tới node con.
- Độ: số node con của một node. Node lá có độ = 0. Độ của cây là độ tối đa của các node trong cây đó.

# Các thuật ngữ

- Khoảng cách: số cạnh dọc đường đi ngắn nhất giữa hai node.
- Độ rộng của một node: số cạnh dọc theo đường đi từ gốc đến node đó.
- Chiều rộng của bậc: số node trong cùng bậc.
- Độ rộng của cây: số node lá trong cây.
- Rừng: tập  $n \geq 0$  cây không liên kết với nhau.
- Kích thước của cây: số lượng node của cây.

# Các thuật ngữ



# Các thao tác với cây

- Liệt kê tất cả các node trong cây.
- Liệt kê một phần của cây.
- Tìm kiếm một phần tử.
- Thêm một phần tử mới vào vị trí xác định trong cây.
- Xóa một phần tử khỏi cây.
- Tỉa cây: cắt bỏ toàn bộ một phần của cây.
- Tìm tổ tiên chung thấp nhất của hai node.
- Duyệt cây theo thứ tự pre-order, in-order, post-order.
  - Pre-order: node cha được duyệt trước node con.
  - Post-order: node con được duyệt trước node cha.
  - In-order: node con bên trái được duyệt trước, sau đó đến node hiện thời và cuối cùng là node con bên phải.

# Biểu diễn cây

- Có nhiều cách khác nhau để biểu diễn cây.
- Cách phổ biến là dùng cấp phát động, mảng.
- Ta cũng có thể sử dụng một danh sách của các danh sách để biểu diễn cây.
- Thông thường trong một cây, node con không chứa liên kết trả tới node cha.



# Nội dung tiếp theo

## Cây nhị phân tìm kiếm