

Cây

Trần Vĩnh Đức

Ngày 18 tháng 1 năm 2017

Nội dung

Cây

Khái niệm cây

Một số tính chất của cây

Ứng dụng

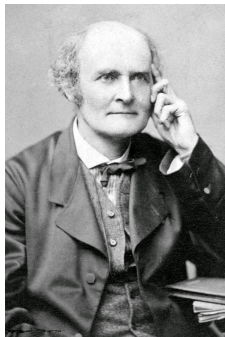
Cây tìm kiếm nhị phân

Cây quyết định

Mã tiền tố

Cây trò chơi

Cây



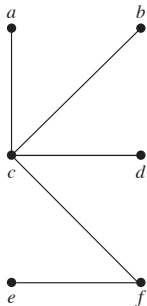
Hình: Arthur Cayley

- ▶ Từ năm 1857, nhà toán học Anh, Arthur Cayley đã dùng cây để xác định những dạng khác nhau của hợp chất hóa học.
- ▶ Cây được dùng để giải nhiều bài toán thuộc nhiều lĩnh vực khác nhau.
- ▶ Trong tin học: cây tìm kiếm, cây quyết định, mã hóa,....

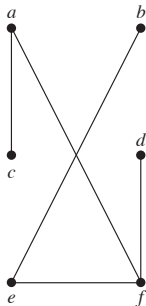
Cây là gì?

Định nghĩa

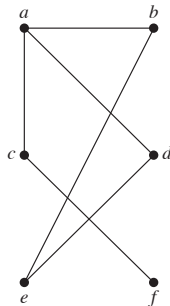
Cây là một đồ thị vô hướng, liên thông và **không** có chu trình đơn.



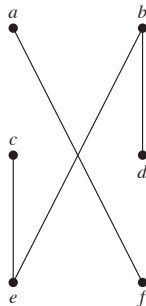
G_1



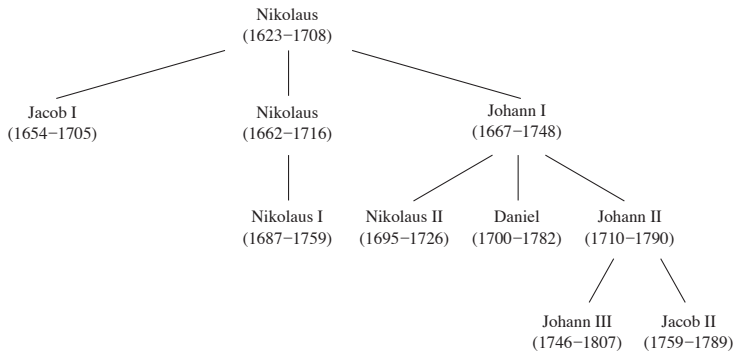
G_2



G_3



G_4

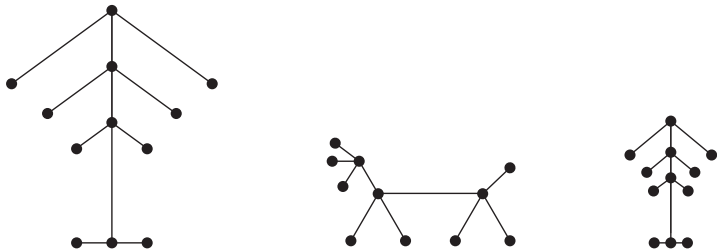


Hình: Cây phả hệ gia đình nhà toán học Bernoulli

Rừng

Định nghĩa

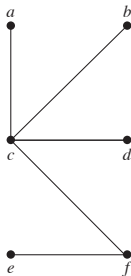
Đồ thị vô hướng **không** có chu trình đơn và **không** liên thông gọi là *rừng*.



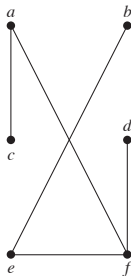
Hình: Rừng với ba thành phần liên thông

Định lý

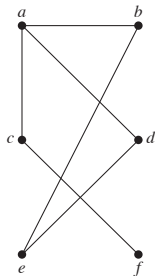
Một đồ thị vô hướng là một cây **nếu và chỉ nếu** giữa mọi cặp đỉnh của nó luôn tồn tại đường đi **duy nhất**.



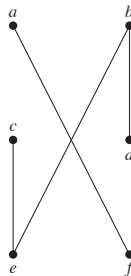
G_1



G_2



G_3

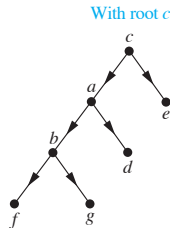
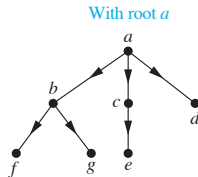
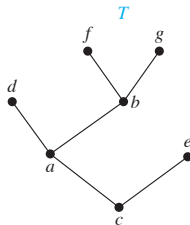


G_4

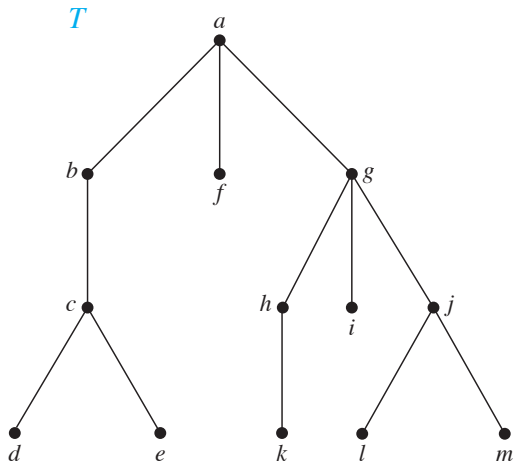
Cây có gốc

Định nghĩa

Cây có gốc là cây có một đỉnh được gọi là gốc và mọi cạnh có hướng từ gốc đi ra.

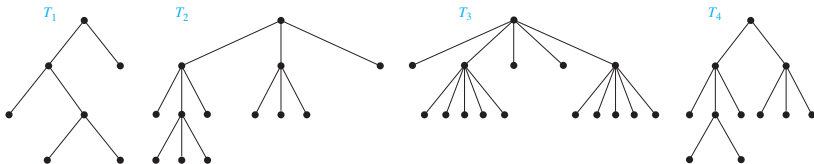


Tổ tiên, Cha, Con, Anh em, Đỉnh trong, Lá

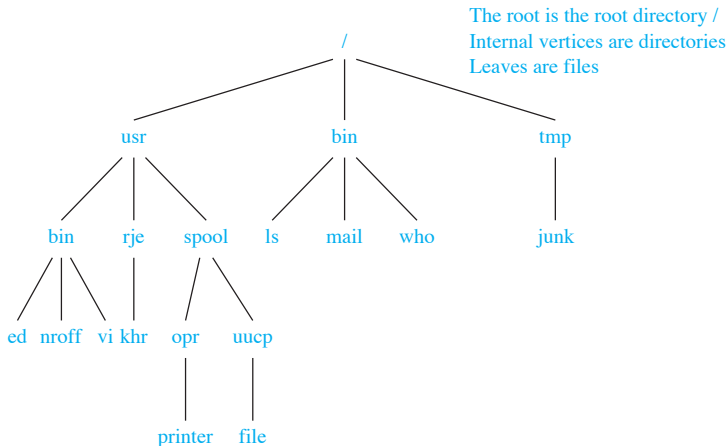


Định nghĩa

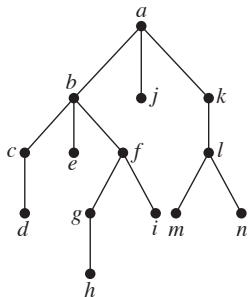
- ▶ Cây có gốc được gọi là cây m -phân nếu tất cả các đỉnh trong của nó có không quá m con.
- ▶ Cây được gọi là m -phân đầy đủ nếu mọi đỉnh trong có đúng m con.
- ▶ Khi $m = 2$ cây được gọi là cây nhị phân.



Mô hình cây



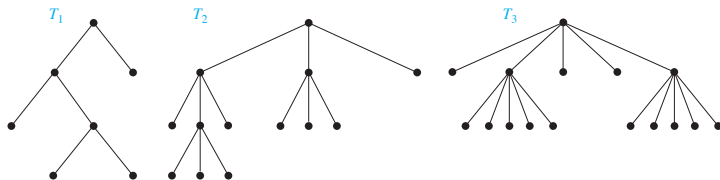
Số cạnh trong cây



Định lý

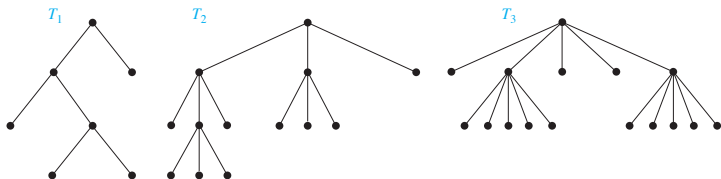
Cây với n đỉnh có đúng $(n - 1)$ cạnh.

Số đỉnh trong và số đỉnh của cây m -phân đầy đủ



Định lý

Cây m -phân đầy đủ với i **đỉnh trong** sẽ có $n = m.i + 1$ **đỉnh**.



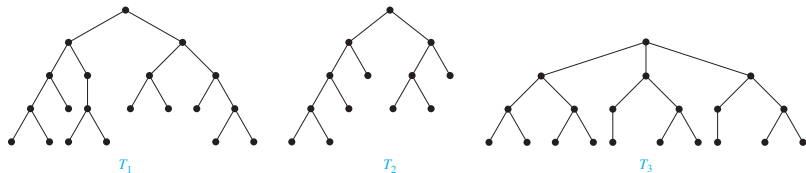
Định lý

Cây m -phân đầy đủ với

- (i) n đỉnh có $i = \frac{n-1}{m}$ đỉnh trong và $\ell = \frac{(m-1)n+1}{m}$ lá.
- (ii) i đỉnh trong có $n = mi + 1$ đỉnh và $\ell = (m-1)i + 1$ lá.
- (iii) ℓ lá có $n = \frac{m\ell-1}{m-1}$ đỉnh và $i = \frac{\ell-1}{m-1}$ đỉnh trong.

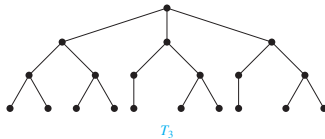
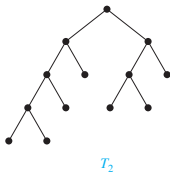
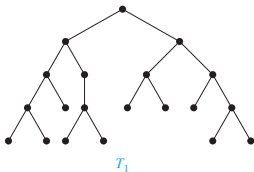
Độ cao

- ▶ **Mức** của đỉnh v là độ dài đường đi duy nhất từ gốc tới v .
- ▶ Mức của gốc là 0.
- ▶ **Độ cao** của cây là mức cao nhất của tất cả các đỉnh.
- ▶ Cây m -phân có gốc và độ cao h được gọi là **cân đối** nếu tất cả các lá đều ở mức h hoặc $(h - 1)$.



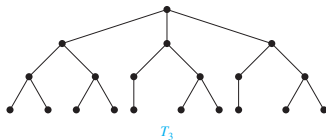
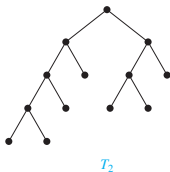
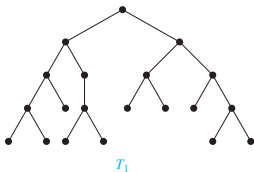
Định lý

Có nhiều nhất m^h lá trong cây m -phân với độ cao h .



Hệ quả

- ▶ Cây m -phân có độ cao h và có ℓ lá, thì $h \geq \lceil \log_m \ell \rceil$
- ▶ Cây m -phân đầy đủ và cân đối có $h = \lceil \log_m \ell \rceil$



Nội dung

Cây

Khái niệm cây

Một số tính chất của cây

Ứng dụng

Cây tìm kiếm nhị phân

Cây quyết định

Mã tiền tố

Cây trò chơi

Cây tìm kiếm nhị phân

- ▶ Để tìm kiếm hiệu quả một phần tử trong danh sách được sắp.
- ▶ Mỗi đỉnh v của cây nhị phân có một khóa thỏa mãn:
- ▶ *Khóa v nhỏ hơn khóa của mọi con bên phải và lớn hơn khóa của mọi con bên trái.*

Ví dụ

Xây dựng cây tìm kiếm nhị phân cho các từ: *mathematics*, *physics*, *geography*, *zoology*, *meteorology*, *geology*, *psychology*, và *chemistry*.

<p>mathematics</p>	<p>mathematics</p> <p>physics</p> <p>physics > mathematics</p>	<p>mathematics</p> <p>geography physics</p> <p>geography < mathematics</p>	<p>mathematics</p> <p>geography physics</p> <p>zoology</p> <p>zoology > mathematics zoology > physics</p>
<p>mathematics</p> <p>geography physics</p> <p>meteorology zoology</p> <p>meteorology > mathematics meteorology < physics</p>	<p>mathematics</p> <p>geography physics</p> <p>geology meteorology zoology</p> <p>geology < mathematics geology > geography</p>	<p>mathematics</p> <p>geography physics</p> <p>geology meteorology zoology</p> <p>psychology</p> <p>psychology > mathematics psychology > physics psychology < zoology</p>	<p>mathematics</p> <p>geography physics</p> <p>geology meteorology zoology</p> <p>chemistry psychology</p> <p>chemistry < mathematics chemistry < geography</p>

Thêm hoặc Tìm một phần tử trong cây

input : Cây tìm kiếm nhị phân T , phần tử x

output: Vị trí của x trong T

$v =$ gốc của T ;

/* Đỉnh không có trong T sẽ có giá trị bằng NULL

*/

while $v \neq \text{NULL}$ và $\text{label}(v) \neq x$ **do**

if $x < \text{label}(v)$ **then**

if con bên trái của $v \neq \text{NULL}$ **then**

$v =$ con bên trái của v ;

end

else

 Thêm đỉnh mới là con trái của v và đặt $v = \text{NULL}$;

end

end

else

if con bên phải của $v \neq \text{NULL}$ **then**

$v =$ con bên phải của v ;

end

else

 Thêm đỉnh mới là con phải của v và đặt $v = \text{NULL}$;

end

end

end

Thêm hoặc Tìm một phần tử trong cây (tiếp)

```
if gốc của  $T = NULL$  then
```

```
    | Thêm đỉnh  $v$  và cây và gán cho nó nhãn là  $x$ ;
```

```
end
```

```
else
```

```
    | if  $v = NULL$  hoặc  $label(v) \neq x$  then
```

```
        | Gán nhãn cho đỉnh mới là  $x$  và đặt  $v$  là đỉnh mới này;
```

```
    end
```

```
end
```

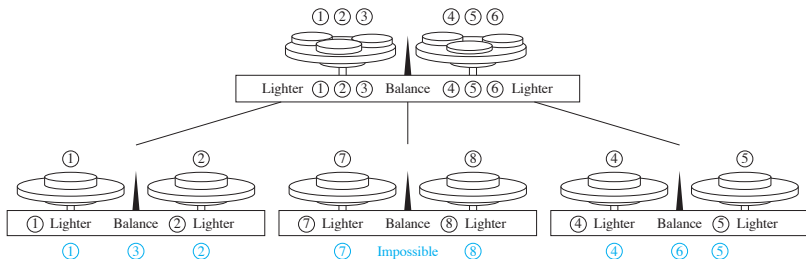
```
/*  $v$  là vị trí của  $x$ 
```

```
*/
```

Cây quyết định

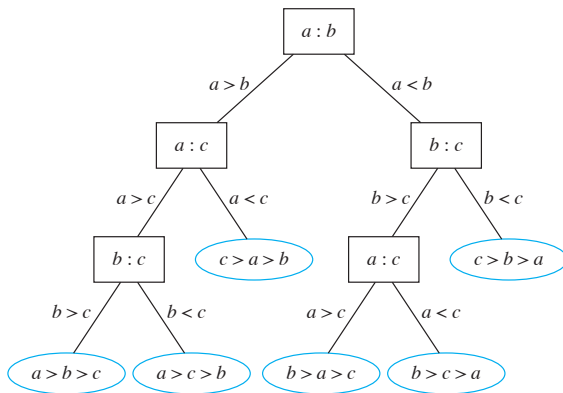
Ví dụ

Có 7 đồng xu có trọng lượng như nhau và một đồng xu giả có trọng lượng nhỏ hơn các đồng khác. Nếu dùng một chiếc cân có hai đĩa thì cần bao nhiêu lần cân để xác định đồng xu giả trong 8 đồng xu?



Ví dụ

Cây quyết định biểu diễn cách sắp xếp ba phần tử khác nhau.



Câu hỏi

Cần ít nhất bao nhiêu phép so sánh nhị nguyên để sắp xếp được dãy gồm n phần tử?

Định lý

Thuật toán sắp xếp dãy với phép so sánh nhị nguyên cần ít nhất $\lceil \log n! \rceil$ phép so sánh.

Chú ý: $\log n! \in \Theta(n \log n)$.

Bài toán mã hóa

- ▶ Mã hóa chữ cái tiếng Anh (không phân biệt chữ hoa hay thường) bằng dãy nhị phân.
- ▶ Có 26 chữ cái, vậy có thể dùng 5 bit để biểu diễn mỗi chữ ($2^4 < 26 < 2^5$).
- ▶ Tổng số bit dùng để mã hóa dữ liệu bằng 5 lần số ký tự trong văn bản.
- ▶ Liệu có thể có một lược đồ mã hóa sao cho khi dữ liệu được mã hóa thì chỉ cần một số bit ít hơn không?

Mã độ dài thay đổi

- ▶ Dùng các xâu nhị phân độ dài khác nhau để mã hóa các chữ cái tiếng Anh.
- ▶ Chữ cái xuất hiện thường xuyên hơn sẽ được mã bằng các xâu nhị phân ngắn hơn.
- ▶ Vấn đề: Làm thế nào xác định được mỗi chữ bắt đầu và kết thúc ở đâu trong xâu mã hóa?

Ví dụ

Cách mã hóa

$$e \rightarrow 0, \quad a \rightarrow 01, \quad t \rightarrow 10$$

gây ra nhập nhằng khi giải mã

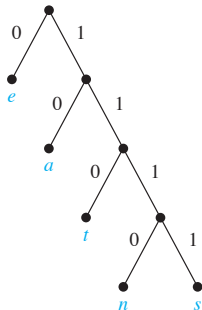
$$010 \rightarrow et$$

$$010 \rightarrow ae$$

Mã tiền tố

Định nghĩa

Mã tiền tố là tập xâu thỏa mãn **không** có xâu nào là khúc đầu của xâu khác.



$e \rightarrow 0,$ $a \rightarrow 10,$
 $t \rightarrow 110,$ $n \rightarrow 1110,$
 $s \rightarrow 1111$

Hãy giải mã xâu được mã hóa bởi

11111011100

Mã Huffman

input : Các ký hiệu a_i với các tần suất w_i với $i = 1, 2, \dots, n$

output: Mã Huffman của các ký hiệu a_i

F = rừng với n cây có gốc, mỗi cây gồm một đỉnh a_i được gán trọng số w_i ;

while F chưa là cây **do**

Thay cây có gốc T và T' có trọng số nhỏ nhất từ F với $w(T) \geq w(T')$ bằng cây có gốc mới với T là cây con trái và T' là cây con phải. Nhãn của cạnh mới nối với T là 0 và cạnh nối với T' là 1 ;

Gán trọng số của cây mới bằng $w(T) + w(T')$;

end

/* mã Huffman của ký tự a_i là dãy các nhãn của các cạnh trên đường đi duy nhất từ gốc tới đỉnh a_i */

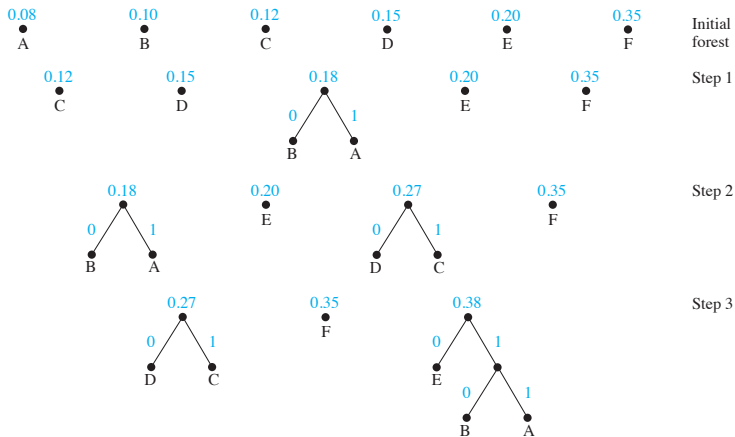
Ví dụ

Dùng mã Huffman để mã các ký tự sau với tần suất

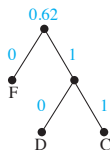
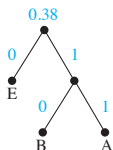
A	B	C	D	E	F
0.08	0.10	0.12	0.15	0.20	0.35

Số bit trung bình dùng để mã hóa một ký tự là bao nhiêu?

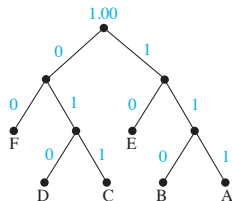
Ví dụ (tiếp)



Ví dụ (tiếp)



Step 4



Step 5

Số bit trung bình để mã hóa một ký tự là

$$3 \times 0.08 + 3 \times 0.10 + 3 \times 0.12 + 3 \times 0.15 + 2 \times 0.20 + 2 \times 0.35 = 2.45$$

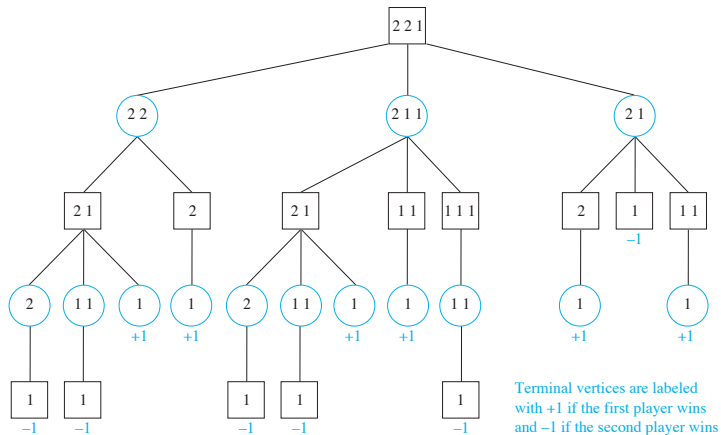
Cây trò chơi

- ▶ Trong một số trò chơi như cờ ca-rô, bốc sỏi, cờ đam và cờ vua, người chơi sẽ đi một nước và biết các nước đi trước đây của đối thủ và không có yếu tố may mắn xen vào.
- ▶ Đỉnh của cây biểu diễn các vị trí (trạng thái) của trò chơi có thể đạt được
- ▶ Cạnh biểu diễn các nước đi hợp lệ giữa các vị trí này.

Trò chơi bốc sỏi

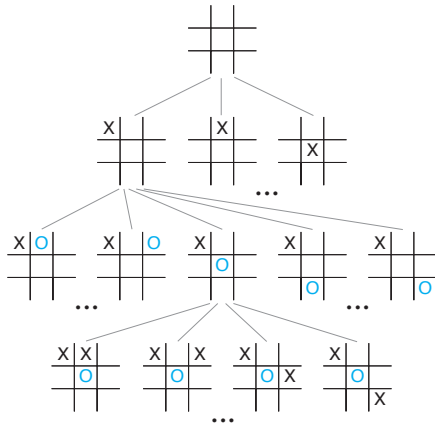
- ▶ Ta có *một số* đồng sỏi.
- ▶ Hai người chơi đến lượt mình sẽ bốc một hoặc nhiều viên sỏi từ *một* trong các đồng sỏi. *Không* được bốc tất cả các viên sỏi.
- ▶ Người nào đến lượt mình không bốc được viên sỏi nào sẽ bị thua.

Trò chơi bốc sỏi



Trò chơi Tic-tac-toe

(a)



(b)

