**Phần 1: Cơ bản Về Giải Thuật**

**Bài 1: giới thiệu về giải thuật**

**Thuật toán Euclid**

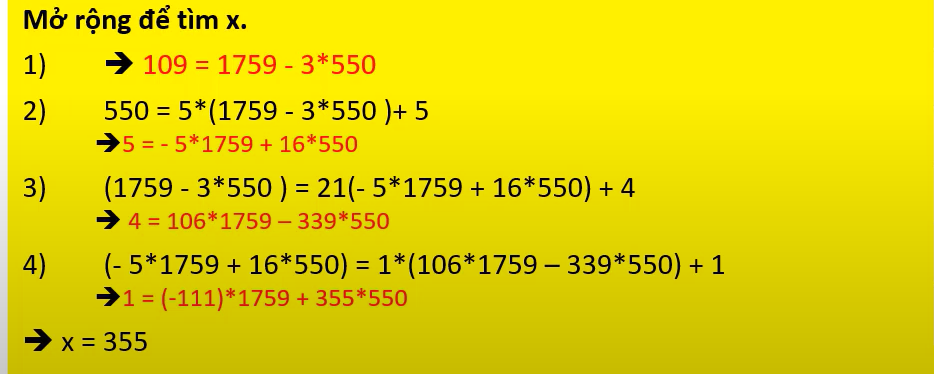
Giả sử cần tìm x sao cho

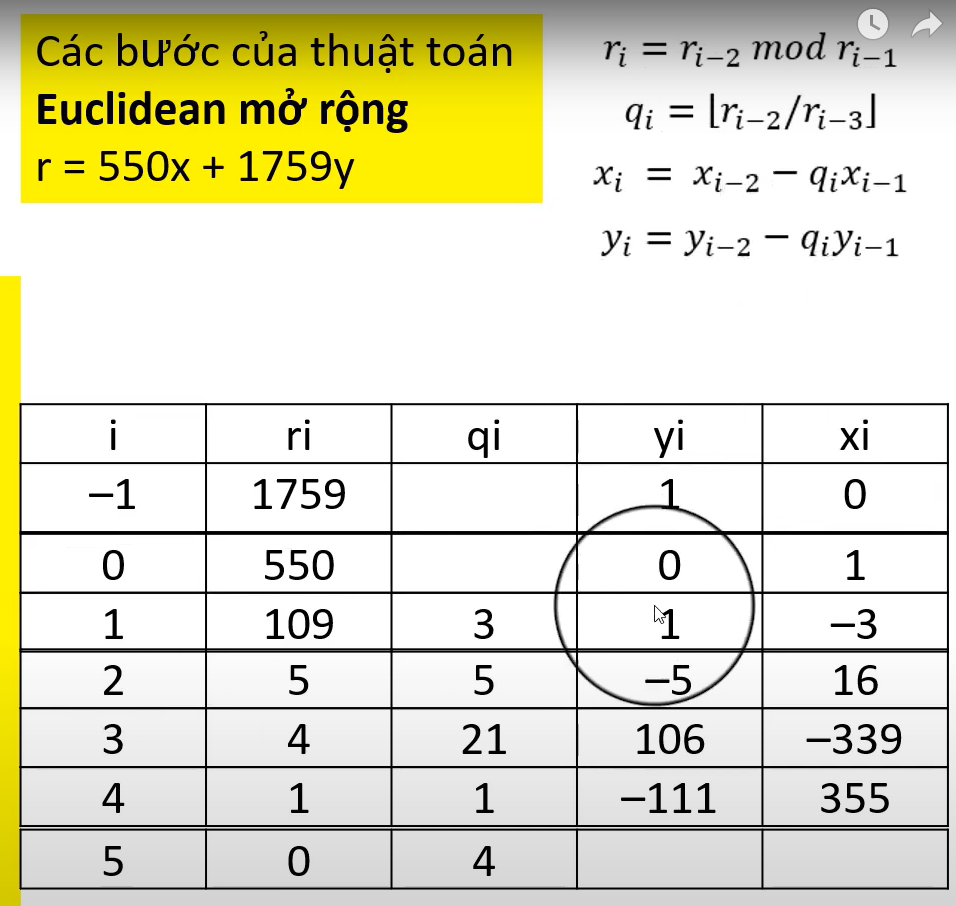
550x = 1 (mod 1759), hay nói cách khác là cần tìm biểu diễn

550x + 1759y = 1

Ta sẽ tìm USCLN:

* Biểu diễn số lớn qua số nhỏ: **1759** = 3 \* **550** + 109
* Biểu diễn số nhỏ qua số dư: **550** = 5 + 109 + 5
* Biểu diễn **109** qua số dư của biểu thức ở trên:: **109** = 21 \* 5 + 4
* Tiếp tục: 5 = 4 \* 1 + 1
* Tiếp tục: 4 = 4 \* 1 + 0





Thuật toán Euclid dùng để xác định các hệ số trong đẳng thức Bezout

Bổ đề Bezout: nếu **d** là ước số chung lớn nhất của hai số nguyên **a** và **b**, thì sẽ tồn tại hai số nguyên **x** và **y** sao cho:

d = a\*x + b \*y

**--------------- --------------- --------------- ---------------**

**Thuật toán nổi bọt**

package com.example.sort**;**import java.util.Scanner**;**public class BubbleSort {  
 static void swap(int[] arr**,** int i**,** int j) {  
 var temp = arr[i]**;** arr[i] = arr[j]**;** arr[j] = temp**;** }  
  
 public static void main(String[] args) {  
 Scanner sc = new Scanner(System.*in*)**;** System.*out*.print("Enter the number of elements you want to store: ")**;** int n = sc.nextInt()**;** int[] array = new int[n]**;** System.*out*.println("Input your array from keyboard:")**;** for (int i = **0;** i < n**;** i++) {  
 array[i] = sc.nextInt()**;** }  
  
 for (int i = **0;** i < array.length**;** i++) {  
 for (int j = array.length - **1;** j > i**;** j--) {  
 if (array[j] < array[j - **1**]) {  
 *swap*(array**,** j**,** j- **1**)**;** }  
 }  
 }  
  
 System.*out*.println("New array after sorted")**;** for (int j : array) {  
 System.*out*.println(j)**;** }  
 }  
}

**--------------- --------------- --------------- ---------------**

**Độ chính xác của thuật toán**

**--------------- --------------- --------------- ---------------**

**Giới thiệu về cấu trúc dữ liệu**

Trong khoa hoc máy tính, cấu trúc dữ liệu là một cách tổ chức dữ liệu trong máy tính sao cho nó có thể được sử dụng một cách hiệu quả cho các thuật toán tác động đến nó

Trong việc thiết kế chương trình, việc chọn cấu trúc dữ liệu là vấn đề quan trọng, kinh nghiệm cho thấy khó khăn của việc triển khai chương trình, chất lượng và hiệu năng của kết quả cuối cùng phụ thuộc rất nhiều vào việc chọn CTDL phù hợp

**Quiz**

**1 A**

**2 A**

**3 A + B**

**4 B + D**

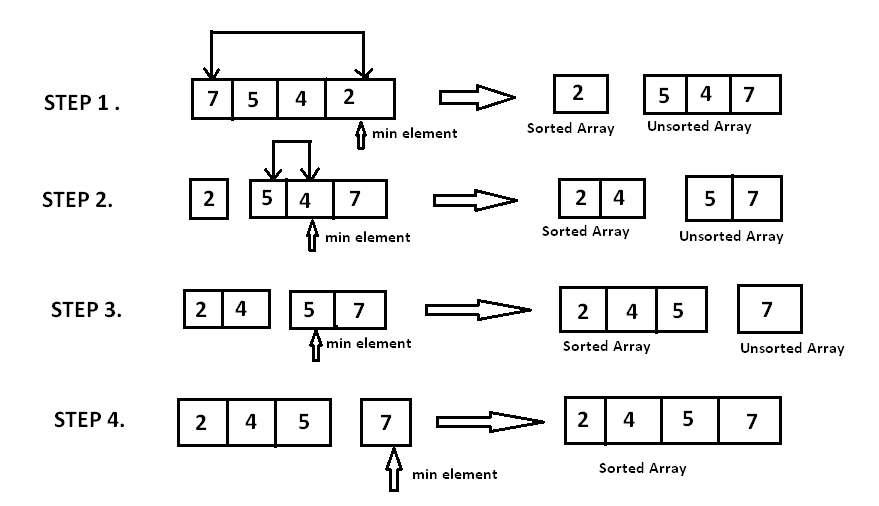
**5 A + B + C**

**--------------- --------------- --------------- --------------- ---------------**

**Thuật toán tìm kiếm và sắp xếp cơ bản**

**Sắp sếp chọn (selection sort)**

Danh sách chứa các phần tử sẽ được chia làm hai phần, phần bên trái là phần được sắp xếp (sort list) và phần bên phải là phần chưa được sắp xếp (unsorted list). Phần tử nhỏ nhất trong list sẽ được chọn và tráo đổi với vị trí đầu tiên, tiếp đến là phần tử nhỏ thứ 2 sẽ được đặt ở vi jtris thử 2 …

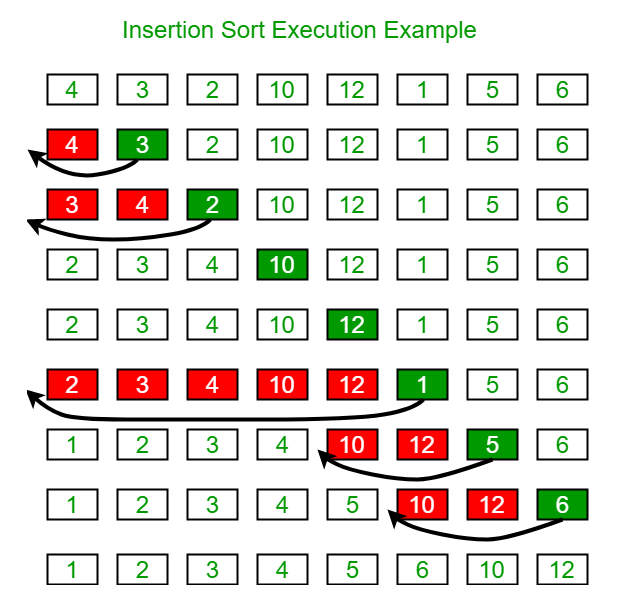


**--------------- --------------- --------------- ---------------**

**Thuật toán chèn (Insertion sort)**

Sắp xếp chèn là thuật toán sắp xếp khi đã có một dãy đã có thử tự, chèn thêm 1 phần tử vào vị trí thích hợp của dãy số đã sắp xếp sao cho dẫy số đó vẫn là dãy có thứ tự.

\



**--------------- --------------- --------------- ---------------**

**Sắp xếp tại chỗ**

**In place (tại chỗ) sorting algorithms** là những thuật toán mà khi chúng ta sắp xếp dữ liệu trong mảng mà không cần sử dụng thêm bộ nhớ bổ sung, điều này có nghĩa là chúng ta sẽ không cần sử dụng các biến tạm để đổi vị trí các element, như vậy tức là có một vùng bộ nhớ đã được sử dụng ngoài dữ liệu lưu trữ của mảng. Các implementations của chúng ta là IN PLACE, có điều nếu chúng ta sử dụng một lượng không đổi bộ nhớ thêm, như trong ví dụ của chúng ta, chúng ta chỉ sử dụng tạm thời một biến – đó có thể đã được định nghĩa một lần bên ngoài vòng lặp và tái sử dụng, việc sắp xếp này là in place

**--------------- --------------- --------------- ---------------**

**Sắp xếp với Stable và Unstable**

Khi sắp xếp các phần tử trong 1 mảng có rất nhiều phần tử có giá trị sắp xếp giống nhau, vậy sau khi sắp xếp mà trình tự của các phần tử **có giá trị giống nhau không thay đổi** thì gọi là stable, ngược lại là unstable

**--------------- --------------- --------------- --------------- --------------- --------------- --------------- --------------- --------------- --------------- --------------- ---------------**

**Phần 2: cấu trúc dữ liệu tuyến tính**

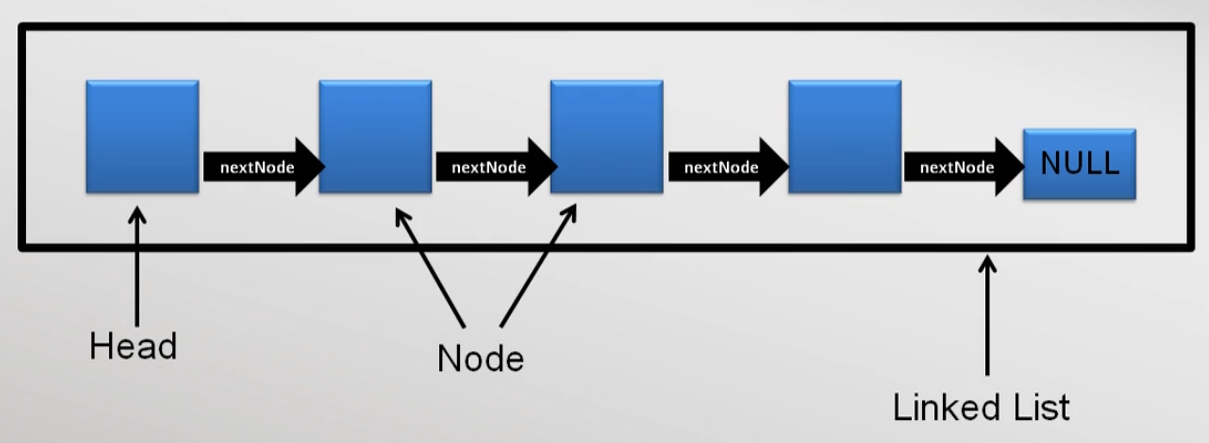
**Bài 3: danh sách dữ liệu Link – List**

**Giới thiệu về danh sách móc nối đơn**

**Danh sách móc nối (link list):** là một kiểu cấu trúc dữ liệu bao gồm một nhóm các node tạo thành một chuỗi. Thông thường mỗi nút gồm dữ liệu (data) ở mút đó và tham chiếu (reference) đến nút kế tiếp trong chuỗi (next). Kiểu cấu trúc dữ liệu này gọi là referential data structure, khác với kiểu array vốn là sequential data structure.

Danh sách móc nối có 3 kiểu chính:

* Danh sách móc nối đơn
* Danh sách móc nối kép
* Danh sách móc nối vòng.



**--------------- --------------- --------------- --------------- --------------- ---------------**

**Bài 4: ngăn xếp (stack) và hàng đợi (queue)**

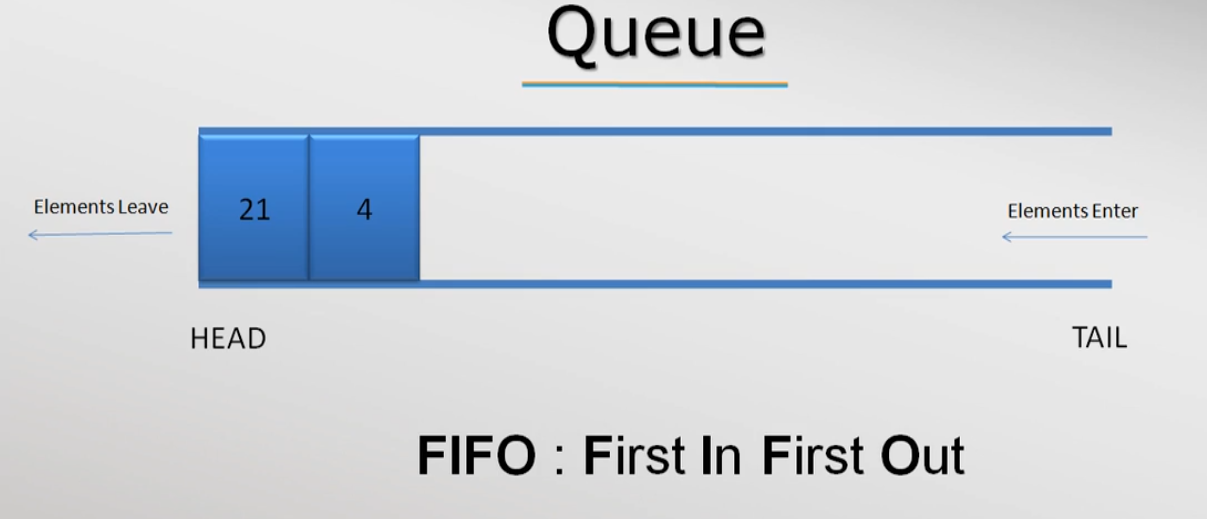
**Stack:** là một cấu trúc dữ liệu trừu tượng hoạt động theo nguyên lý “vào sau ra trước” (last in first out).

Trong **Stack,** các đối tượng có thể đc bổ sung vào stack bất kỳ lúc nào nhưng chỉ được phép lấy ra phần tử ở đỉnh của stack.

**Queue**: Một ngăn xếp là một cấu trúc dữ liệu dạng thùng chưa (container) của các phần tử (thường được gọi là các node) và có hai phép toán cơ bản: push và pop. Push là bổ sung một phần tử vào đỉnh (top) của ngăn xếp, nghĩa là bổ sung vào sau các phần tử đãcó trông ngăn xếp, pop là giải phóng, trả về phần tử đứng ở đỉnh của ngăn xếp. Đối với Queue sẽ là First In First Out,

Đối với **Queue** ta có các method như sau:

* Enqueue: thêm một element vào từ tail
* Dequeue: xóa phần tử head ra khỏi queue và return ra phần tử đó.
* Peek: trả về phần tử ở head mà không remove phần tử đó ra khỏi queue.



**--------------- --------------- --------------- --------------- --------------- ---------------**

**Bài 5: Đệ quy**

**Giới thiệu về đệ quy**

Đệ quy là phương pháp dùng trong các chương trình máy tính, trong đó một hàm tự gọi đến chính nó. Hay nói cách khác, một khái niệm X được định nghĩa theo đệ quy nếu như trong định nghĩa X có sử dụng ngay chính khải niệm X.

Một hàm đệ quy căn bản gồm hai phần:

1. Phần cơ sở (hay còn gọi là tính dừng của đệ quy): chứa các tác động của hàm hoặc thủ tục với một số giá trị cụ thể ban đầu của tham số
2. Phần đệ quy: định nghĩa tác động cần thực hiện cho giá trị hiện thời của các tham số bằng các tác động đã được định nghĩa trước đây với kích thước tham số nhỏ hơn

**Đệ quy đuôi (tail recursion)**

**Tail Recurion** là hàm đệ quy nhưng việc gọi đệ quy được thực hiện ở câu lệnh cuối cùng của hàm, do đó không có bất kỳ câu lệnh nào bị block (câu lệnh chưa được thực hiện)

Ngược lại với **Tail Recurion** tức là hàm mà đệ quy được thực hiện không phải ở vị trí cuối cùng 🡺 có câu lệnh sẽ bị block. Tuy nhiên Java không support việc sử dụng phương pháp này.

**Bài toán tháp Hà Nội**

Có một bộ đĩa gồm các kích thước khác nhau, có lỗ ở giữa, nằm xuyên trên 3 cái cọc. Bài toán rằng phải sắp xếp các đĩa theo trật tự kích thước vào một cọc sao cho các đĩa tạo thành dạng hình nón với đĩa nhỏ nhất nằm trên cùng. Quy tắc chuyển đĩa:

1. Mỗi lần chỉ được chuyển một đĩa
2. Một đĩa chỉ được đăt lên trên đĩa lớn hơn.

**--------------- --------------- --------------- --------------- --------------- --------------- --------------- --------------- --------------- --------------- --------------- ---------------**

**Phần 3: cấu trúc dữ liệu phi tuyến tính**

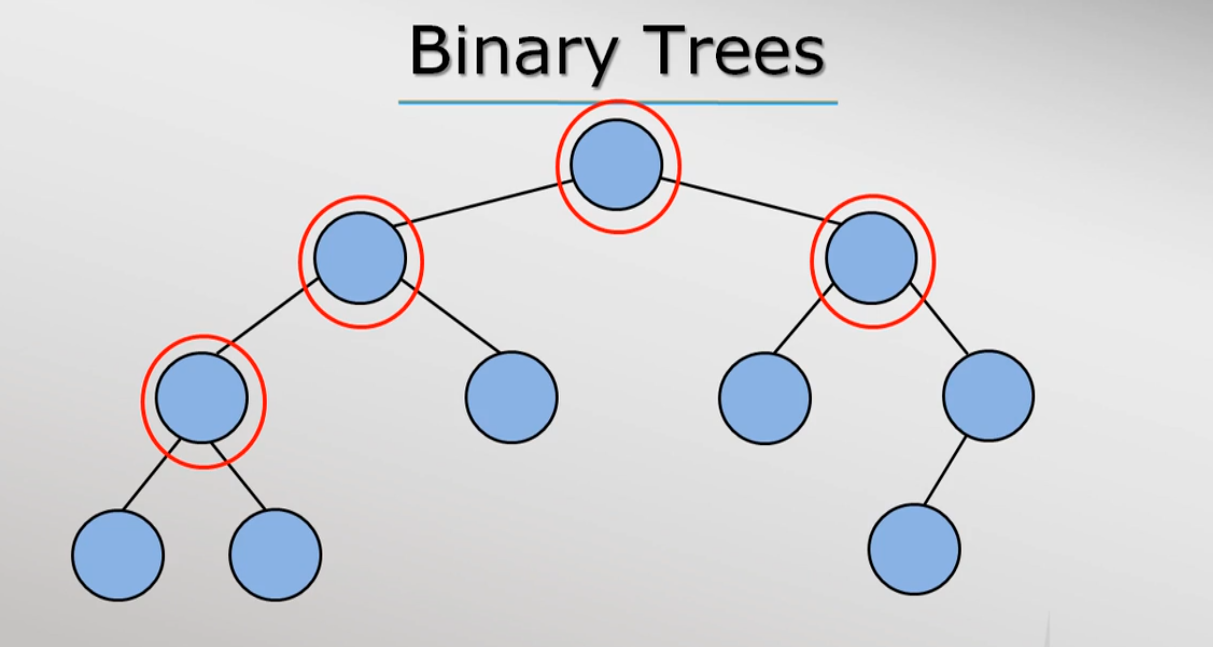
**Bài 6: Cấu trúc dữ liệu cây**

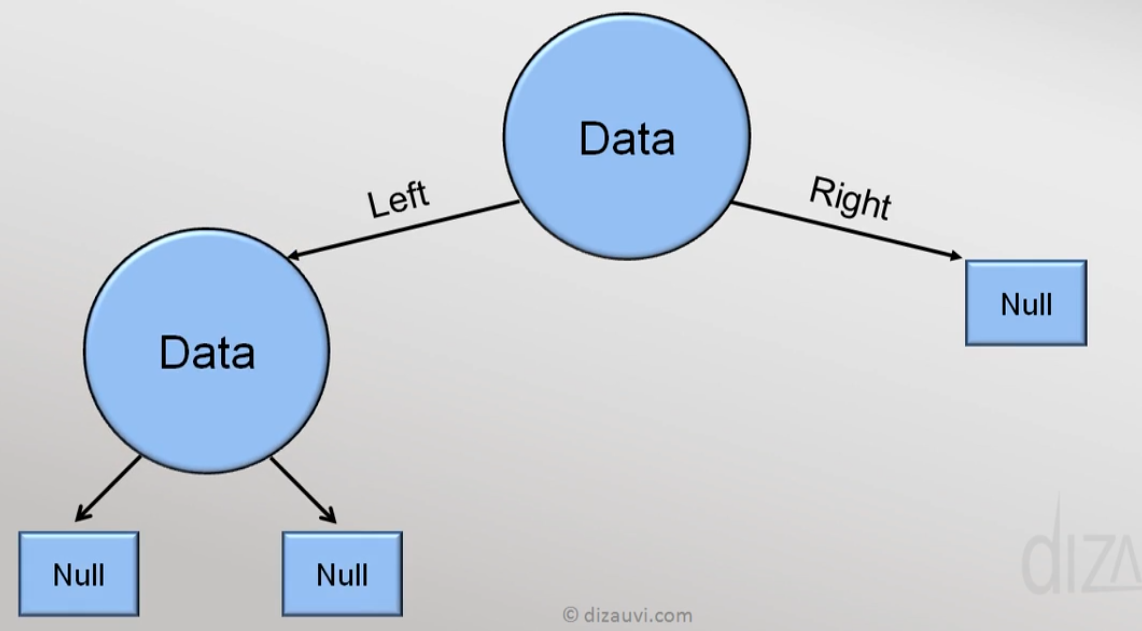
**Giới thiệu cấu trúc dữ liệu cây**

Cấu trúc dữ liệu cây (tree) là một abstract data type được dùng để lưu trữ cấu trúc phân cấp. do đó Tree không phải là cấu trúc dữ liệu tuyến tính. Nó cũng lưu trữ data trong các **node.** Lấy nguồn cảm hứng từ các cây gia phả vậy.

**Cây nhị phân**

Là cấu trúc **Tree** nhưng mỗi node chỉ có 2 node con, mỗi node con sẽ được biểu diễn bằng một con trỏ **left** và **right.** Cây nhị phân hội tụ đủ các lợi ích của mảng sắp xếp (sorted array) và danh sách liên kết (linked list). Chẳng hạn tìm kiếm sẽ nhanh hơn như rtrong sorted array, và việc chèn hoặc xóa phần tử trong cây cũng nhanh như trong linked list



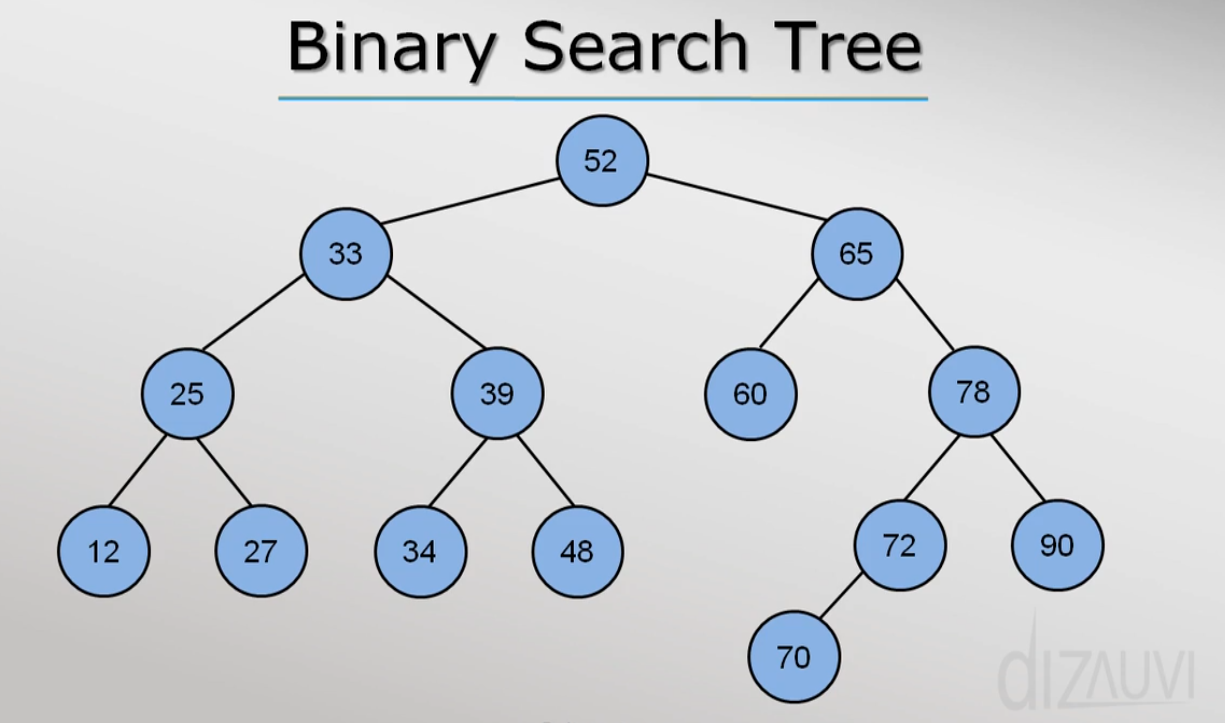


**--------------- --------------- --------------- ---------------**

**Binary Search Tree**

Là một Binary Tree thỏa mãn những điều kiện sau:

* Tất cả các Left Node có value nhỏ hơn Parent Node, Grandparent Node … của nó.
* Tất cả các Right Node có value lớn hơn Parent Node, Grandparent Node … của nó

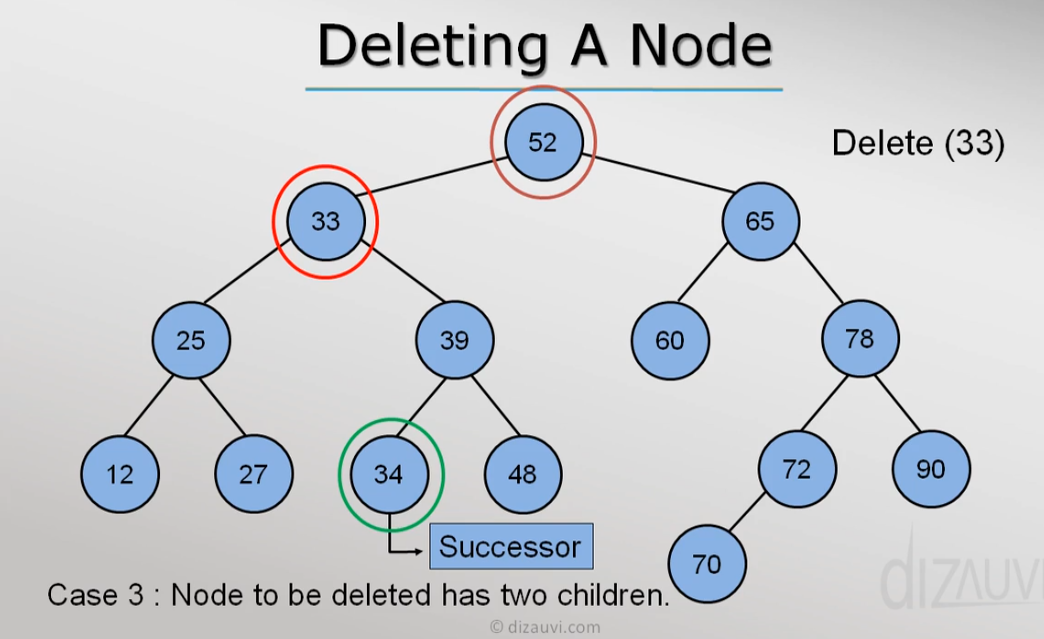


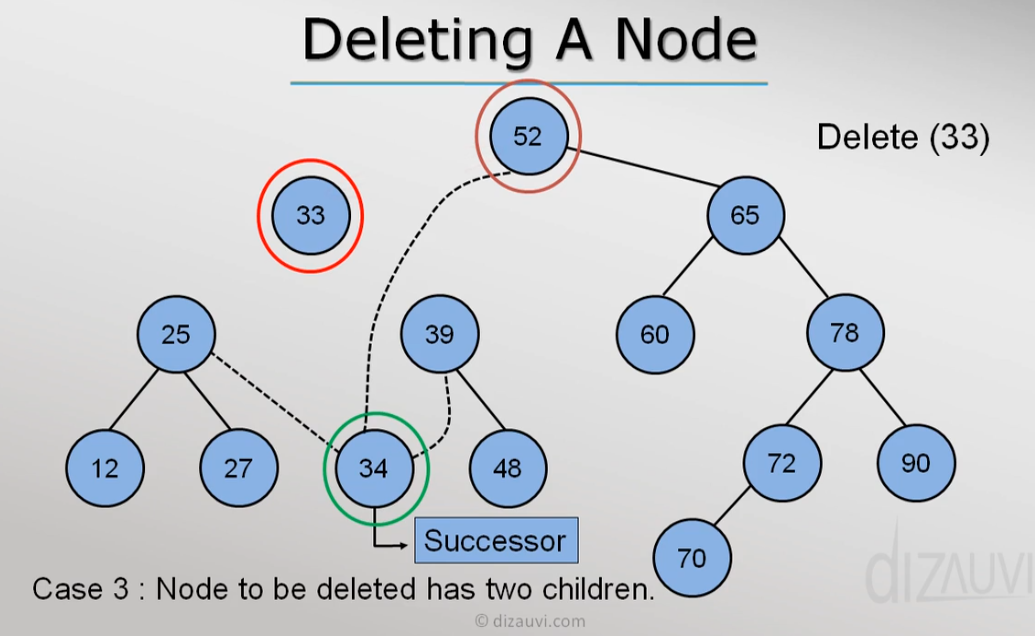
**--------------- --------------- --------------- ---------------**

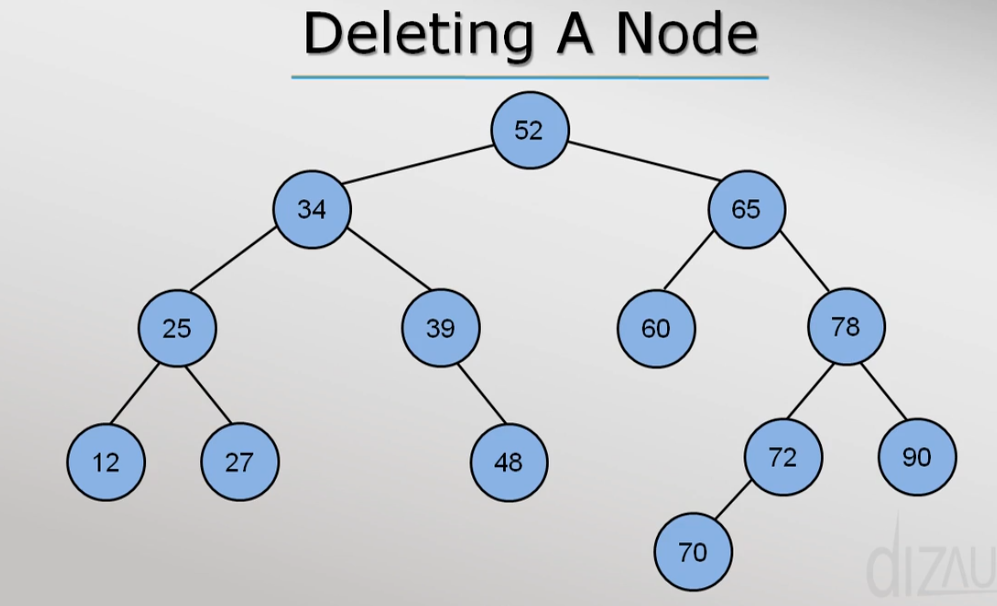
**Xóa một phần tử**

Với case 1 và case 2, khá đơn giản.

Với case 3, là node cần xóa có cả left và right child Node:

****

****

****

Cách xóa trên khá là phức tạp, do đó chúng ta thường sử dụng một phương pháp khác, đó là **Soft Delete**

**Soft Delete** là không hẳn xóa node đi, mà chúng ta sẽ đánh dấu là xóa, tuy nhiên mỗi khi chúng ta duyệt qua Tree, hoặc tìm kiếm một item, chúng ta sẽ không quan tâm đến phần tử bị đánh dấu là đã xóa này.

**--------------- --------------- --------------- ---------------**

**Duyệt cây nhị phân**

**In-Order traversal:**

* Duyệt left sub tree
* Visit the root
* Duyệt the right sub tree

**Pre-Orrder Traversal:**

* Visit the Root
* Duyệt sang left sub tree
* Duyệt sang right sub tree

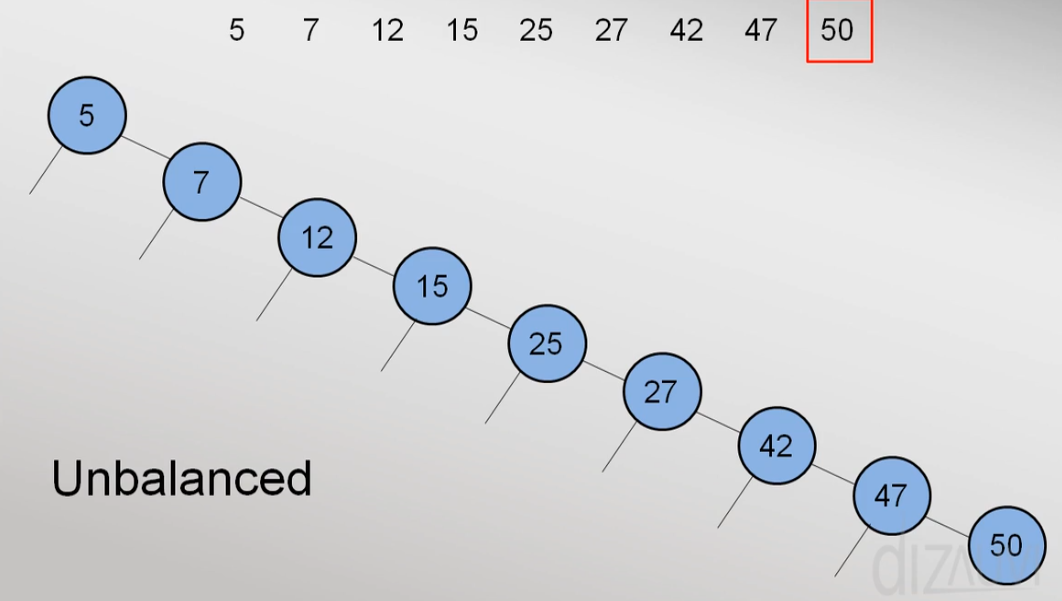
**Post-Order Traversal**

* Duyệt qua left sub tree
* Duyệt tiếp đến right sub tree
* Cuối cùng visit the Root

**--------------- --------------- --------------- ---------------**

**Unbalanced Trees & Balanced Tree**

Gỉa sử chúng ta liên tục thêm các phần tử có giá trị một cách tuần tự (tăng dần), thì khi đó các phần tử được thêm vào sẽ luôn luôn là phần tử right child node đối với parent. => Unbalanced, đây gần giống như là LinkedList, khi đó chúng ta sẽ không dùng được các đặc tính của một TreeNode nữa, khi đó nếu duyệt để tìm kiếm phần thử, ta sẽ có O(n) là độ phức tạp.

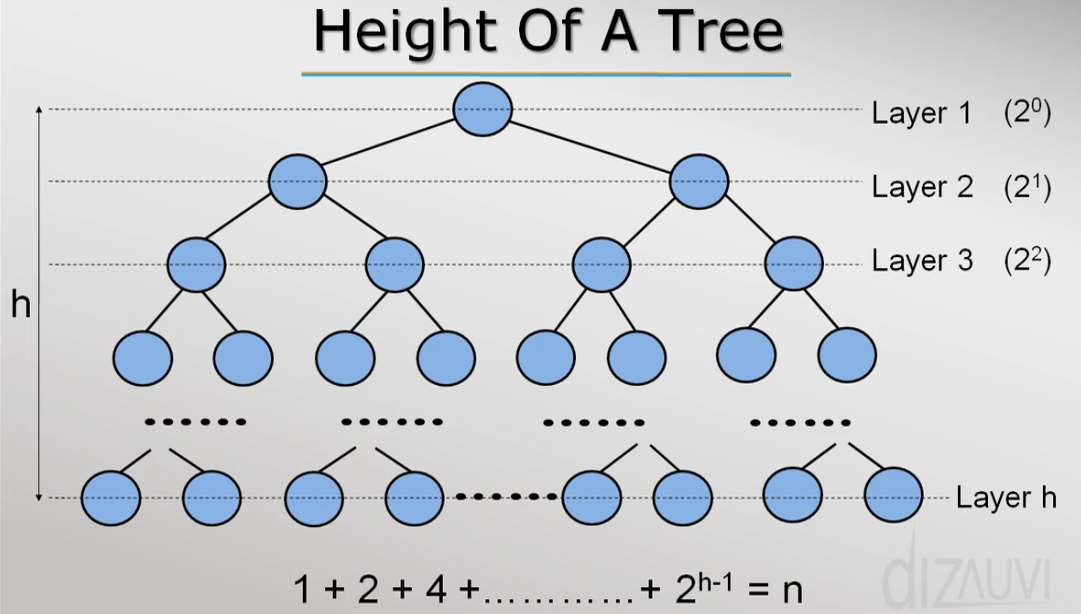


Các operations mà chúng ta thực hiện trên một Tree sẽ đạt hiệu quả tốt nhất trên một Balanced Tree, do đó lý tưởng nhất, chúng ta luôn luôn muốn có một balanced tree, tuy nhiên điều này phụ thuộc vào data mà chúng ta dùng để build nên Tree đó.

**--------------- --------------- --------------- ---------------**

**Height Of a Tree**

Giả sử chúng ta có một Balanced Tree, chúng ta sẽ xác định height của nó.



**--------------- --------------- --------------- ---------------**

**Time Complexity of Operations on Binary Search Tree**

Giả sử chúng ta xem xét trên một Balanced Search Tree, ta có n là số lượng các Node của Tree.

Search (item) = O(log2n), tương tự với Delete, vì chúng ta cũng phải tìm kiếm Node trước đó và thực hiện 1 vài tác vụ trên Node để thực hiên xóa Node đó đi, việc insert item cũng là tương tự.

**--------------- --------------- --------------- --------------- --------------- ---------------**

**Bài 7: đồ thị**

Là khái niệm về sự biểu diễn của các connections giữa các object, miêu tả rất nhiều sự vật, bối cảnh, các vấn đề trogn cuộc sống cũng như khoa học máy tính.

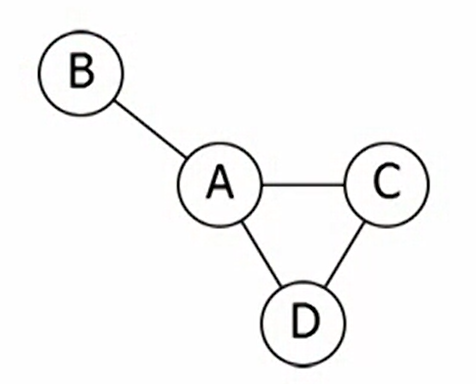
Graph: là tập hợp các các V (of vertices – các đỉnh), và các E (of Edge – các cạnh) sẽ kết nổi các đỉnh với nhau

**--------------- --------------- --------------- ---------------**

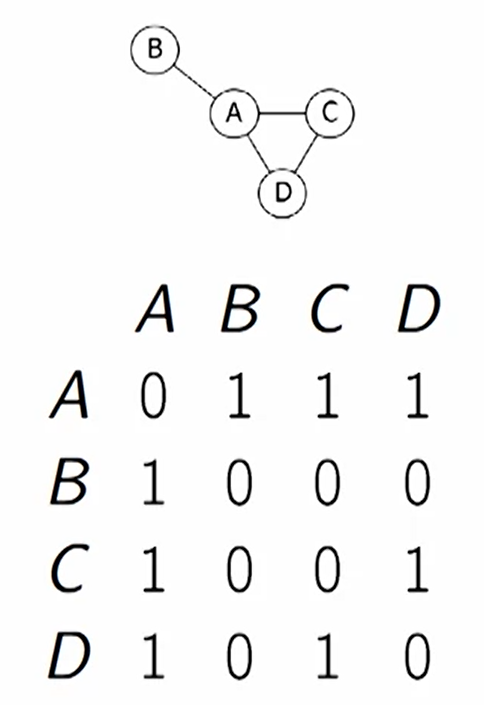
**Biểu diễn đồ thị**

Sử dụng Edge List:

* Chúng ta có các node: A, B, C, D
* Chúng ta có các cạnh: (A, B), (A, C), (A, D), (C, D)



Sử dụng ma trận gần kề: mỗi entries sẽ bằng 1 nếu như 2 edge có thể connect được đến nhau



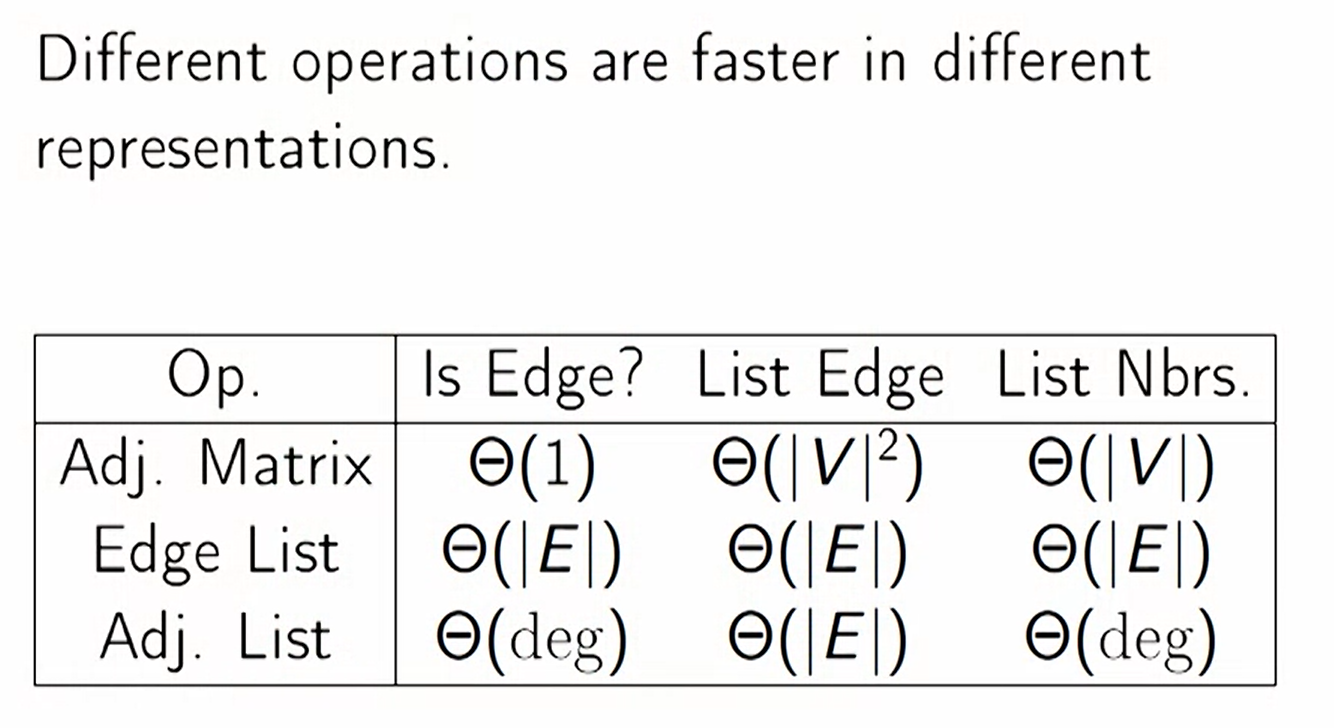
Sử dụng các danh sách liền kề: mỗi một đỉnh sẽ có một list các đỉnh mà nó có connect được đến, do đó ta sẽ có list như sau đối với ma trận ở trên:

A adjacent to B, C, D

B adjacent to A

C adjacent to A, D

D adjacent to A, C



Thời gian chạy của thuật toán Graph:

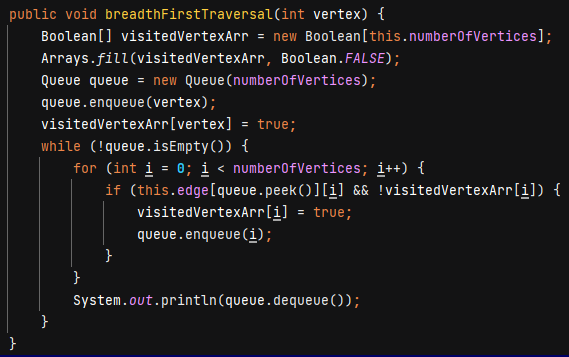
* phụ thuộc vào số lượng các Đỉnh |V| và các Cạnh |E|

**--------------- --------------- --------------- ---------------**

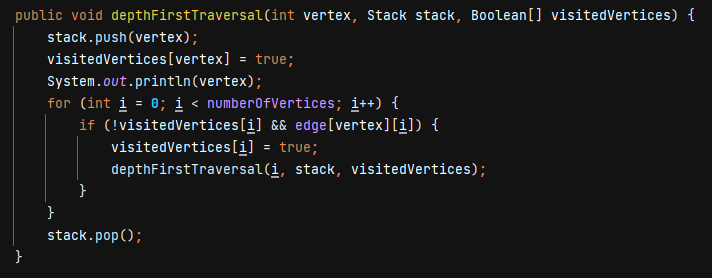
**Duyệt đồ thị theo chiều rộng**

Có 2 phương pháp để duyệt toàn bộ các đỉnh của một đồ thị:

1. Duyệt (tìm kiếm) theo chiều rộng (breadth first search): thuật toán BFS bắt đầu từ đỉnh V (được xác định trước) và lần lượt nhìn các đỉnh liền kề với đỉnh V. Sau đó với mỗi đỉnh trong số đó, thuật toán lại lần lượt nhìn các đỉnh liền kề với nó mà chưa được quan sát trước đó và lặp lại cho đến khi tất cả các đỉnh đều được quan sát.



1. Duyệt theo chiều sâu (depth first search): cũng bắt đầu tại một đỉnh V cho trước, quá trình tìm kiếm được phát triển tới đỉnh con đầu tiên của nút đang tìm kiếm và cứ như thế tiếp tục như thế cho tới khi gặp được đỉnh cần tìm hoặc tới 1 đỉnh không có con, khi đó giải thuật quay lui về đỉnh vừa mới tìm kiếm ở bước trước đó để tìm một khả năng khác.



**Đồ thị vô hướng**

Đồ thị mà trong đó tất cả các cạnh đều vô hướng (không có hướng) thì được gọi là đồ thị vô hướng, mỗi cạnh luôn là một mối quan hệ hai chiều và có thể duyệt theo hai hướng. Đồ thì có hướng là trường hợp ngược lại của đồ thị vô hướng, tức là từ một đỉnh A có thể đến đỉnh B, nhưng điều ngược lại chưa chắc xảy ra.

**Tinh liên thông**

Một đồ thị được gọi là liên thông nếu như mọi cặp đỉnh trong đồ thị đều có đường đi đến nhau, nếu tồn tại một cặp đỉnh không có đường đi đến nhau, ta gọi đồ thị đó là không liên thông.

**Giới thiệu về bài toán Định tuyến nhanh nhất**

Định tuyến là xác định tuyến đường đi theo một điều kiện nhất định

Khi định nghĩa về chiều dài đường đi sẽ có hai cách:

* Chiều dài đường đi chính là số các cạnh mà nó đi qua
* Chiều dài đường đi chính là tổng số trọng số của các đoạn mà nó đi qua, áp dụng với đồ thị có trọng số

**Thuật toán Dijkstra**

Khi đi từ một đỉnh A đến đỉnh B trong đồ thị, chúng ta có thể có rất nhiều đường đi, thường ra sẽ chọn đường đi có độ dài ngắn nhất. Thuật toán Dijkstra là một trong số những thuật toán có thể giải được bài toán này.

Ý tưởng của thuật toán này là lưu một tập hợp các vertices mà khoảng cách đã được đặt chính xác (known region)

Đỉnh đầu tiên được added vào R là S (source)

Sau mỗi lần lặp chúng ta sẽ lấy 1 đỉnh không thuộc R với dist-value nhỏ nhất, gán nó vào R, và tiếp tục relax những cạnh tiếp theo.