Nêu bài toán

• Cho hai xâu  
X = (x1,x2,…,xm) và  
Y = (y1,y2,…,yn)  
• Hãy tìm xâu con chung dài nhất của hai dãy X và Y

Ý tưởng thuật toán

• Phân rã:  
– m: chiều dài xâu X,

n: chiều dài xâu Y  
– Với mỗi 0≤ i ≤ m và 0 ≤ j ≤ n gọi MAX[i, j] là độ dài của  
dãy con chung dài nhất của hai dãy  
Xi=x1x2…xi và Yj =y1y2…yj  
(Qui ước X0 = rỗng, Y0= rỗng)  
– Khi đó MAX[m,n] là chiều dài xâu con chung dài nhất của X và Y.

• Bài toán con: MAX[0,j]=0 j=1..n, MAX[i,0] = 0 i=1..m

Tổng hợp  
• Với i > 0, j > 0 tính MAX[i, j]  
– Nếu xi = yj thì dãy con chung dài nhất của Xi và Yj  
sẽ thu được bằng việc bổ sung xi (cũng là yj) vào dãy con chung dài nhất của hai dãy Xi-1 và Yj-1  
– Nếu xi ≠ yj thì dãy con chung dài nhất của Xi và Yj  
sẽ là dãy con dài hơn trong hai dãy con chung dài nhất của (Xi−1 và Yj) và của (Xi và Yj−1)  
  
MAX[i,j] = MAX[i-1,j-1]+1MAX[i,j] = Max{MAX[i-1,j], MAX[i,j-1]}

Cài đặt

For(int i=0;i<=m;i++){

MAX[i][0]=0;

}

For(int j=0;j<=n;j++){

MAX[0][j]=0;

}

For(int i=1;j<=m;j++){

For(int j=1;j<=n;j++){

If(Xi==Yj) MAX[i][j]=MAX[i-1][j-1]+1;

Else MAX[i][j]=MAX[i][j-1]>MAX[i-1][j]?MAX[i][j-1]:MAX[i-1][j];

}

}

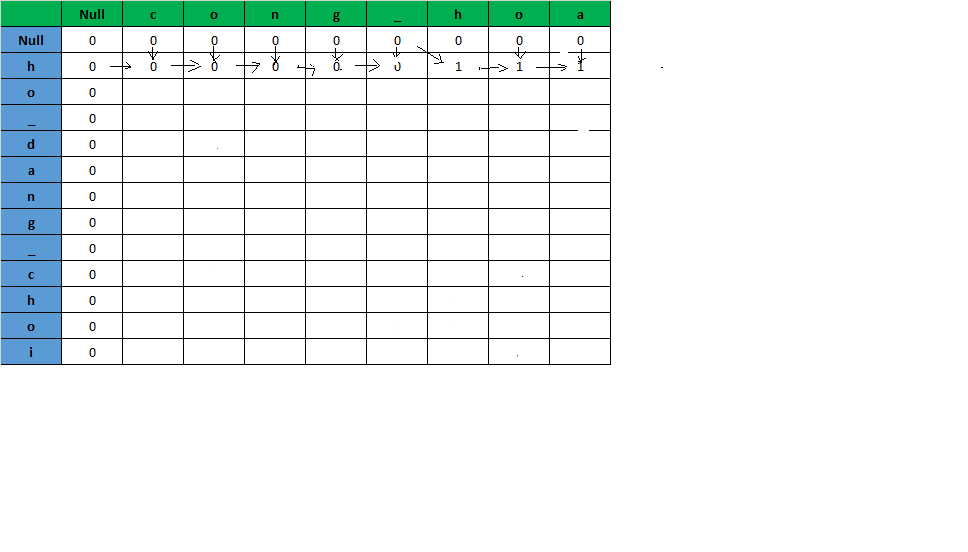
Độ phức tạp

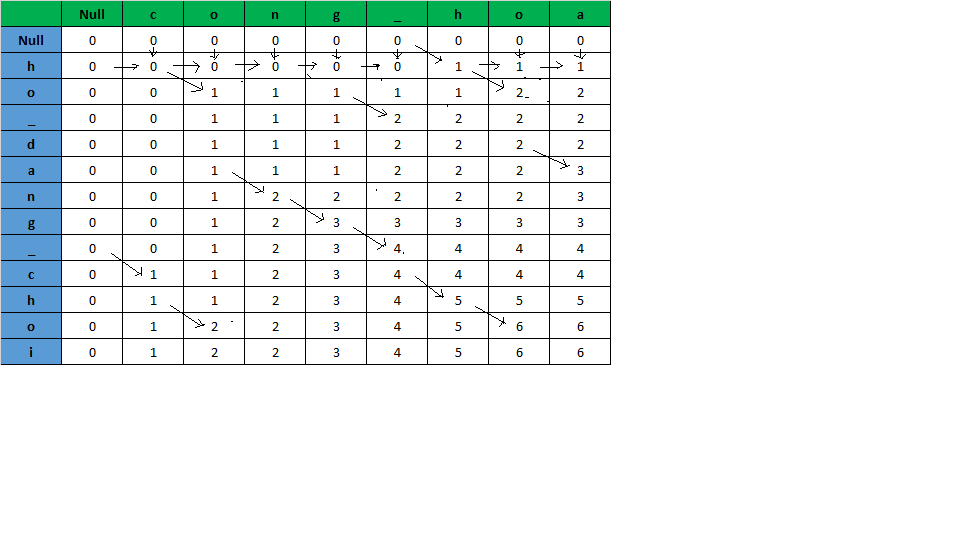
Fn=(m+1) + (n+1) + 6\*m\*n=O(m\*n)

Ví dụ:

Chuỗi 1: cong hoa

Chuỗi 2: ho dang choi





Kết thúc: MAX[m][n]=6

* Độ dài xâu con chung lớn nhất là: 6

Đưa ra thêm 1 cách giải quyết bài toán: Cây hậu tố tổng quát

Cây hậu tố T của xâu S gồm m kí tự là một cây có hướng, có gốc chứa đúng m là được đánh số từ 1 tới m. Mỗi một nút bên trong cây, khác với nút gốc, có ít nhất hai con và mỗi cạnh được gán nhãn bởi một xâu con không rỗng của S. Không có hai cạnh nào đi ra từ một nút mà có nhãn của cạnh bắt đầu bằng cùng một kí tự. Đặc điểm quan trọng của cây hậu tố đó là, với bất kì lá i nào, nhãn của cạnh trên đường từ gốc tới lá i chính là hậu tố của S bắt đầu từ vị trí i, tứclà S[i..m].[1]

Nhãn đường của nút v (L(v)) là xâu con được gán nhãn cho cạnh bắt đầu từ gốc và kết thúc tại nút đó, hay chính là nhãn của đường từ gốc của cây tới nút đó. [2]

Độ sâu của một nút: Với mỗi nút v của một cây hậu tố, độ sâu của nút v là độ dài của L(v) (tức là số kí tự trên đường). [2]

Bài toán LCS của 2 xâu S1 và S2 với cây hậu tố tổng quát được thực hiện như sau:

* Bước 1: Xây dựng cây hậu tố tổng quát đối với S1 và S2
* Bước 2: Đánh dấu mỗi nút bên trong v của cây bằng số 1 (hoặc 2) nếu cây con dưới v chứa một lá đối với một hậu tố của S1 (hoặc S2)
* Bước 3: Duyệt cây tìm các nút được đánh số cả 1 và 2 và chọn nút u nào mà có độ sâu lớn nhất và khi đó L(u)chính là xâu con dài nhất