

# Xây dựng DFA từ biểu thức chính quy

Hoàng Văn Tuấn

Email: [tuanhoang.97dx@gmail.com](mailto:tuanhoang.97dx@gmail.com)

# Xây dựng DFA từ biểu thức chính quy

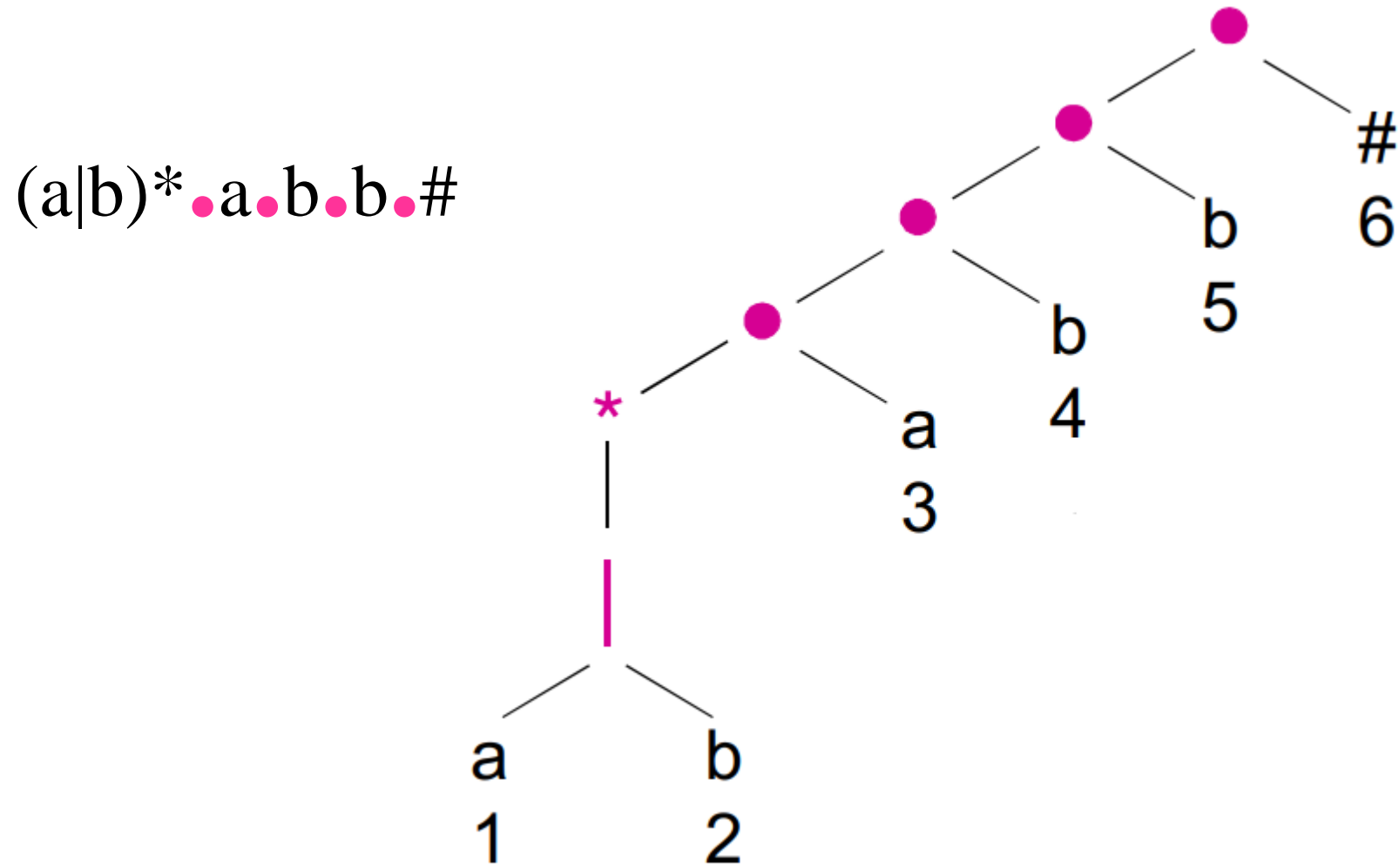
- Input: Biểu thức chính quy  $r$
- Output: Một DFA  $D$  biểu diễn cho  $L(r)$
- Phương pháp:
  - Xây dựng cây phân tích cú pháp cho biểu thức chính quy  $r$  gia tố  $r\#$
  - Tính các giá trị của hàm nullable, firstpos, lastpos và followpos cho các nút trên cây cú pháp
  - Xây dựng tập các trạng thái  $Dstates$ , hàm chuyển trạng thái  $Dtran$  cho DFA

# 1. Cây phân tích cú pháp

- Các nút lá là các ký hiệu của  $r\#$ 
  - Mỗi lý hiệu kết thúc (gồm cả  $\#$ ) sẽ được đánh thứ tự
- Các nút trong là các toán tử:
  - Ghép (Cat Node) ●
  - Hợp (Or Node) |
  - Bao đóng (Star Node) \*

# 1. Cây phân tích cú pháp

- Cây cú pháp cho biểu thức chính quy  $(a|b)^*abb\#$

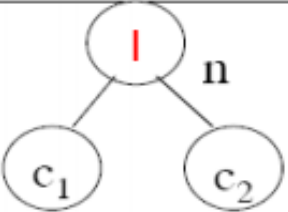
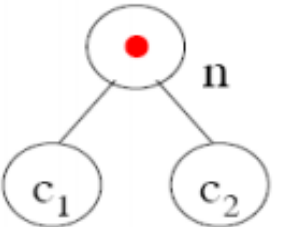
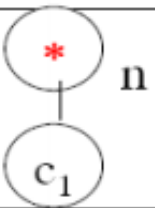


## 2. Tính nullable, firstpos, lastpos, followpos

- Nullable(n): Nhận giá trị true nếu cây con với gốc tại n sinh ra ngôn ngữ gồm cả chuỗi rỗng. Ngược lại, thì nó nhận giá trị false
- Firstpos(n): Tập các vị trí ứng với những ký tự đầu tiên của chuỗi sinh ra bởi cây con có gốc tại n
- Lastpos(n): Tập các vị trí ứng với ký tự kết thúc của chuỗi sinh ra bởi cây con có gốc tại n
- Followpos(n): Tập các vị trí mà có thể đứng ở sau vị trí p trong biểu thức chính quy gia tổ. Followpos chỉ định nghĩa cho node lá, không cho các node trong.

## 2. Tính nullable, firstpos, lastpos, followpos

- Quy tắc tính nullable, firstpos, lastpos

Nút n	<i>nullable</i> (n)	<i>firstpos</i> (n)	<i>lastpos</i> (n)
n là nút có tên là $\epsilon$	true		
n là nút có tên là vị trí i	false	{i}	{i}
	<i>nullable</i> ( $c_1$ ) <b>or</b> <i>nullable</i> ( $c_2$ )	<i>firstpos</i> ( $c_1$ ) $\cup$ <i>firstpos</i> ( $c_2$ )	<i>lastpos</i> ( $c_1$ ) $\cup$ <i>lastpos</i> ( $c_2$ )
	<i>nullable</i> ( $c_1$ ) <b>and</b> <i>nullable</i> ( $c_2$ )	if <i>nullable</i> ( $c_1$ ) then <i>firstpos</i> ( $c_1$ ) $\cup$ <i>firstpos</i> ( $c_2$ ) else <i>firstpos</i> ( $c_1$ )	if <i>nullable</i> ( $c_2$ ) then <i>lastpos</i> ( $c_1$ ) $\cup$ <i>lastpos</i> ( $c_2$ ) else <i>lastpos</i> ( $c_2$ )
	true	<i>firstpos</i> ( $c_1$ )	<i>lastpos</i> ( $c_1$ )

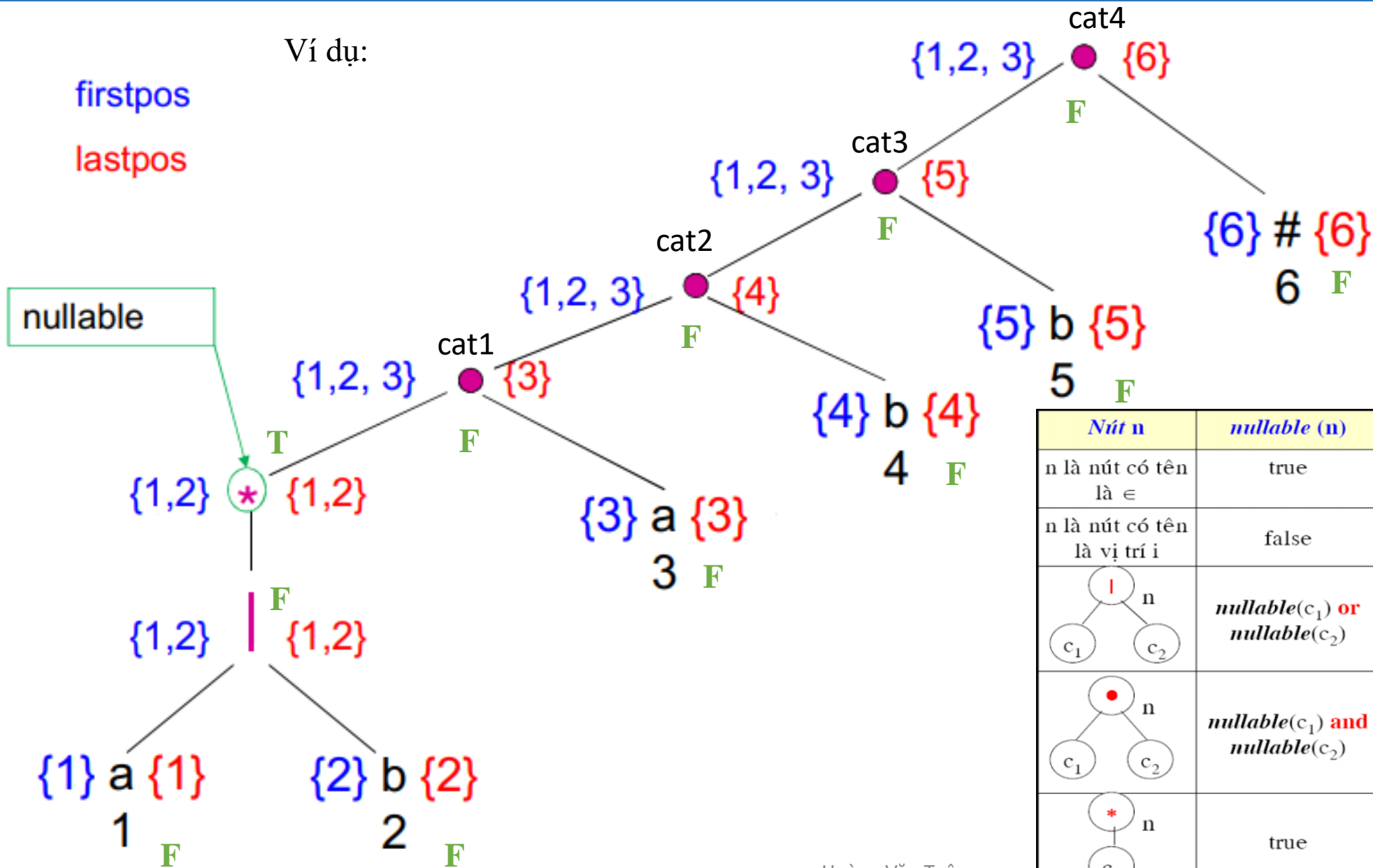
# Bổ trợ về phép toán logic

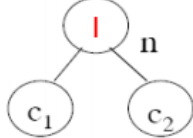
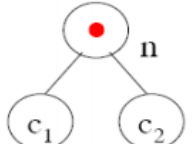
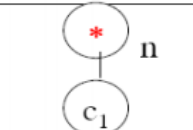
A	B	A <b>and</b> B
False	False	False
False	True	False
True	False	False
True	True	True

A	B	A <b>or</b> B
False	False	False
False	True	True
True	False	True
True	True	True

## 2. Tính nullable, firstpos, lastpos, followpos

Ví dụ:

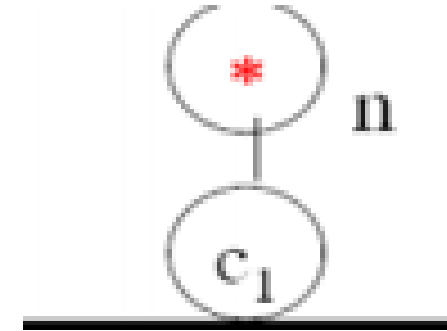
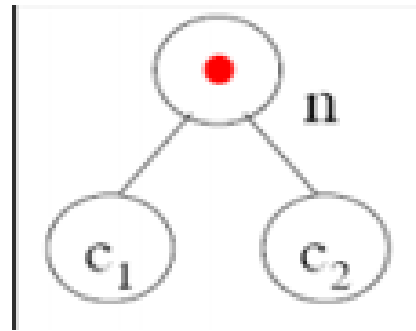


<i>Nút n</i>	<i>nullable (n)</i>	<i>firstpos (n)</i>	<i>lastpos (n)</i>
n là nút có tên là $\epsilon$	true		
n là nút có tên là vị trí i	false	{i}	{i}
	<i>nullable(c<sub>1</sub>) or nullable(c<sub>2</sub>)</i>	<i>firstpos(c<sub>1</sub>) <math>\cup</math> firstpos(c<sub>2</sub>)</i>	<i>lastpos(c<sub>1</sub>) <math>\cup</math> lastpos(c<sub>2</sub>)</i>
	<i>nullable(c<sub>1</sub>) and nullable(c<sub>2</sub>)</i>	<i>if nullable(c<sub>1</sub>) then firstpos(c<sub>1</sub>) <math>\cup</math> firstpos(c<sub>2</sub>) else firstpos(c<sub>1</sub>)</i>	<i>if nullable(c<sub>2</sub>) then lastpos(c<sub>1</sub>) <math>\cup</math> lastpos(c<sub>2</sub>) else lastpos(c<sub>2</sub>)</i>
	true	<i>firstpos(c<sub>1</sub>)</i>	<i>lastpos(c<sub>1</sub>)</i>

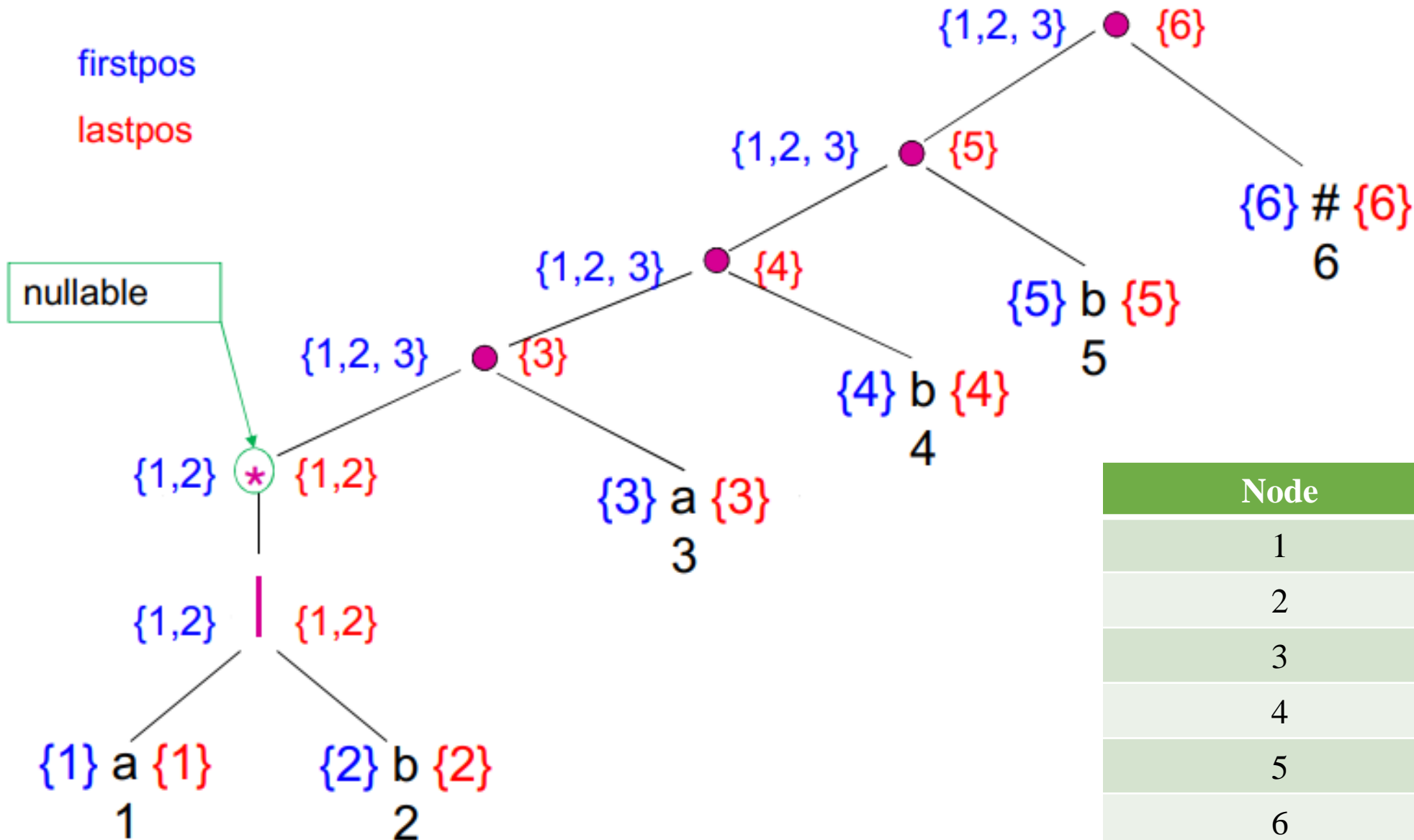


## 2. Tính nullable, firstpos, lastpos, followpos

- Sau khi ta tính nullable, firstpos và lastpos thì followpos được xác định như sau:
  - Nếu nút  $n$  là cat – node (phép ghép) với con bên trái là  $c_1$ , con bên phải là  $c_2$ , và  $i$  là vị trí trong  $\text{lastpos}(c_1)$ , thì tất cả các vị trí trong  $\text{firstpos}(c_2)$  sẽ cho vào  $\text{followpos}(i)$
  - Nếu  $n$  là star – node (bao đóng) và  $i$  là vị trí trong  $\text{lastpos}(n)$  thì tất cả các vị trí trong  $\text{firstpos}(n)$  sẽ cho vào  $\text{followpos}(i)$



## 2. Tính nullable, firstpos, lastpos, followpos



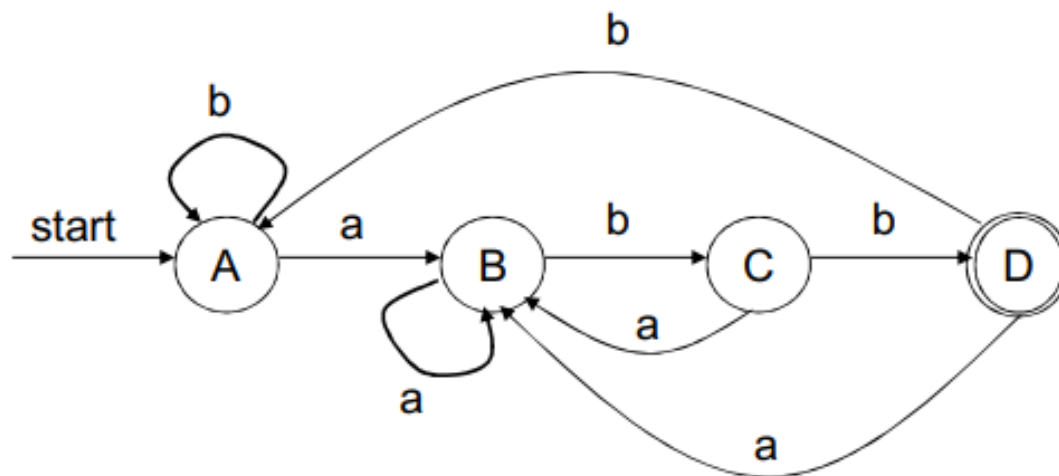
Node	followpos
1	$\{1, 2, 3\}$
2	$\{1, 2, 3\}$
3	$\{4\}$
4	$\{5\}$
5	$\{6\}$
6	-

### 3. Xây dựng $Dstates$ và $Dtran$

```
initialize  $Dstates$  to contain only the unmarked state  $firstpos(n_0)$ ,  
    where  $n_0$  is the root of syntax tree  $T$  for  $(r)\#$ ;  
while ( there is an unmarked state  $S$  in  $Dstates$  ) {  
    mark  $S$ ;  
    for ( each input symbol  $a$  ) {  
        let  $U$  be the union of  $followpos(p)$  for all  $p$   
            in  $S$  that correspond to  $a$ ;  
        if (  $U$  is not in  $Dstates$  )  
            add  $U$  as an unmarked state to  $Dstates$ ;  
         $Dtran[S, a] = U$ ;  
    }  
}
```

## 4. Ví dụ

- Hãy xây dựng DFA từ biểu thức chính quy sau:  $(a|b)^*abb$



Trạng thái	Ký tự đầu vào	
	a	b
A	B	A
B	B	C
C	B	D
D	B	A

Trong đó:

$A = \{1, 2, 3\}$

$B = \{1, 2, 3, 4\}$

$C = \{1, 2, 3, 5\}$

$D = \{1, 2, 3, 6\}$

Thank you for  
watching