

## □ 词法分析概述及自动化需要解决的问题

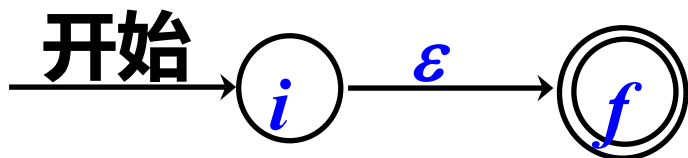
## □ 词法分析器的自动生成

- ❖ 词法单元的描述：正则式
- ❖ 词法单元的识别：转换图
- ❖ 有限自动机：NFA、DFA
- ❖ 正则表达式 → NFA → DFA → 化简的DFA

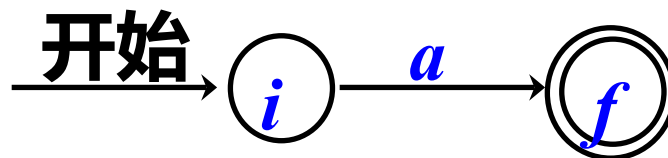


□ 首先构造识别 $\varepsilon$ 和字母表中一个符号 $a$ 的NFA

❖ 重要特点：仅一个接受状态，它没有向外的转换



识别正则表达式 $\varepsilon$ 的  
NFA



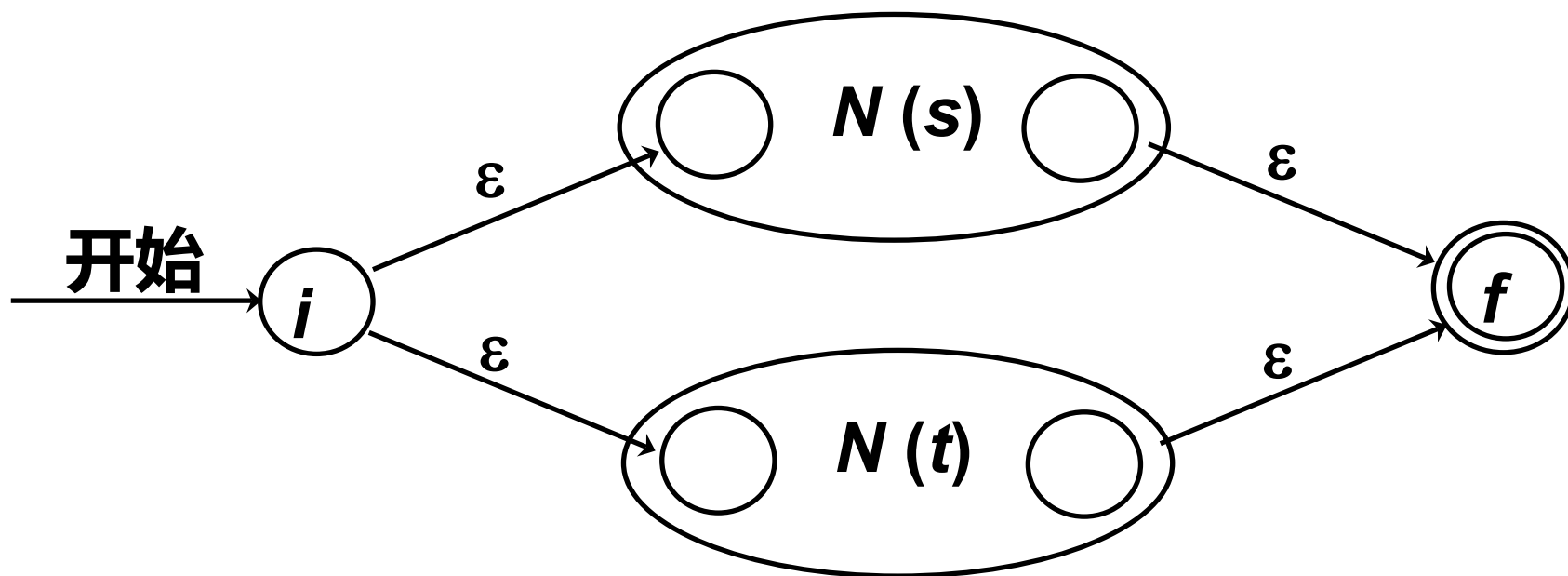
识别正则表达式 $a$ 的  
NFA

□ 对于加括号的正则表达式( $s$ )，其NFA可用 $s$ 的NFA（用 $N(s)$ 表示）代替



## □ 构造识别主算符为选择的正则表达式的NFA

❖ 重要特点：仅一个接受状态，它没有向外的转换

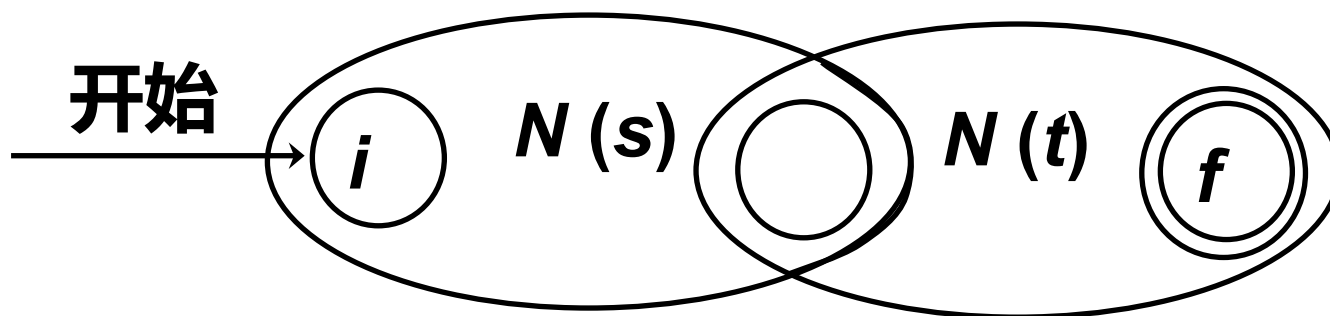


识别正则表达式  $s \mid t$  的NFA



## □ 构造识别主算符为连接的正则表达式的NFA

❖ 重要特点：仅一个接受状态，它没有向外的转换

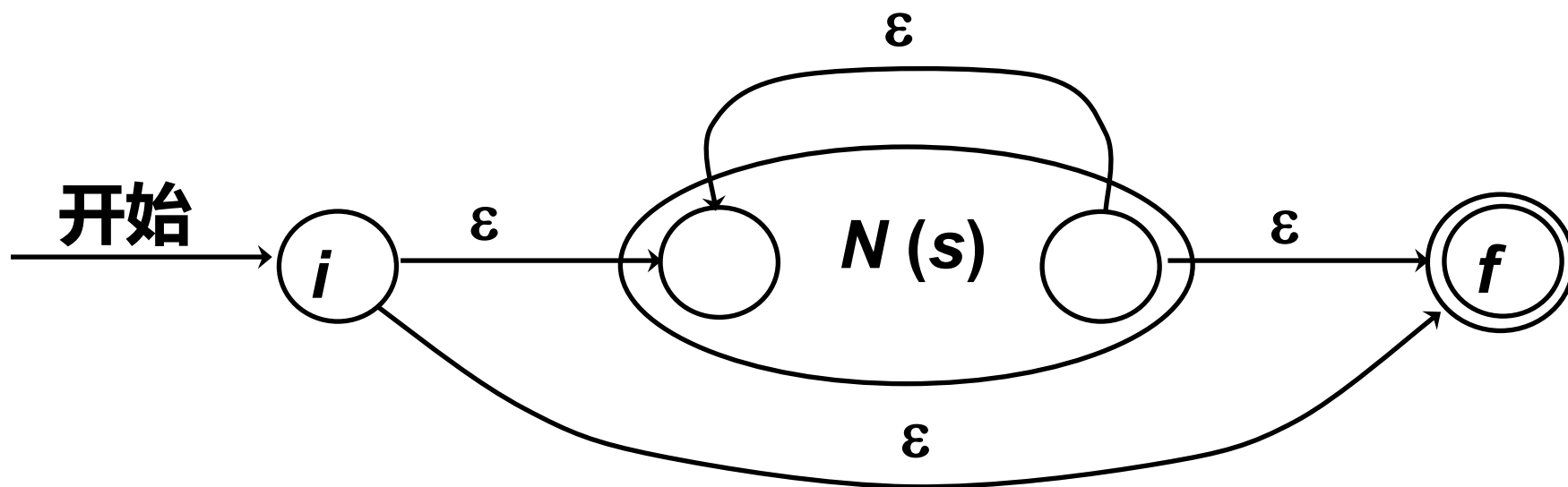


识别正则表达式 $st$  的NFA



## □ 构造识别主算符为闭包的正则表达式的NFA

❖ 重要特点：仅一个接受状态，它没有向外的转换

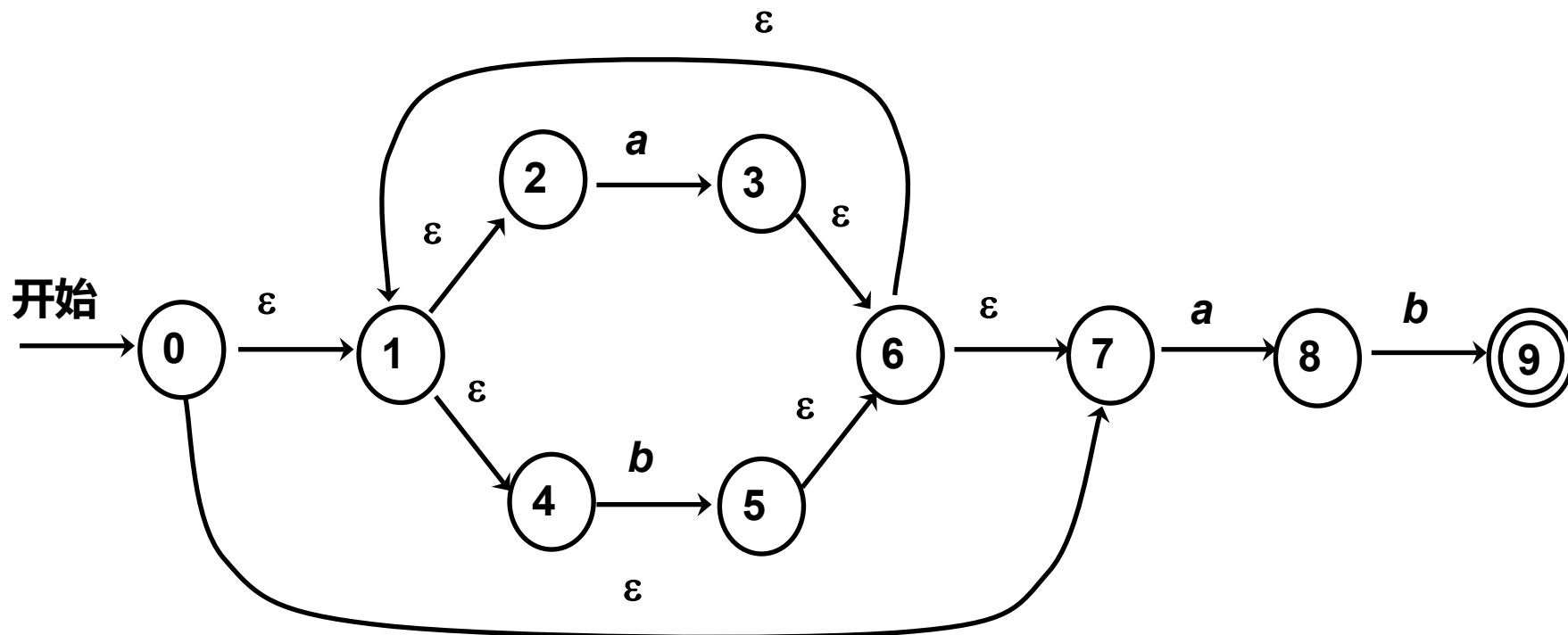


识别正则表达式 $s^*$ 的NFA



## □ 由本方法产生的NFA具有下列性质：

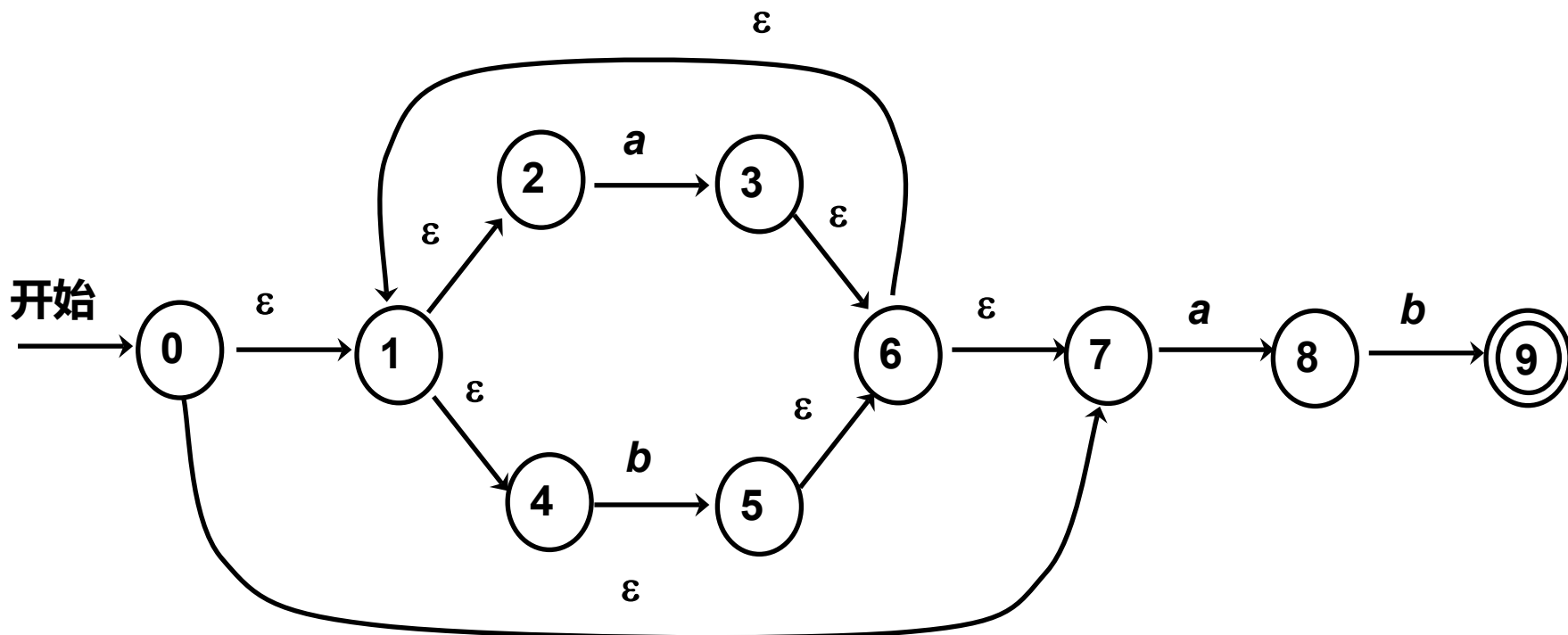
- ❖  $N(r)$  的状态数最多是 $r$ 中符号和算符总数的两倍
- ❖  $N(r)$  只有一个接受状态，接受状态没有向外的转换





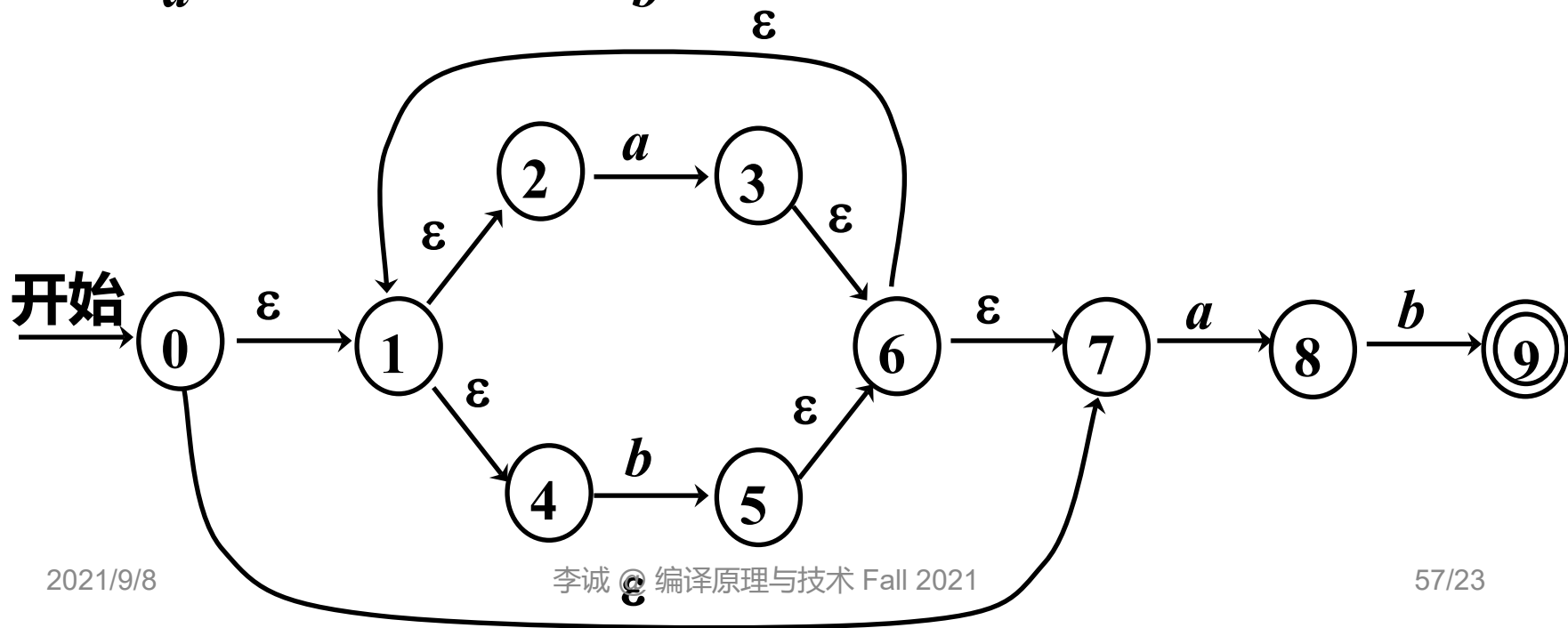
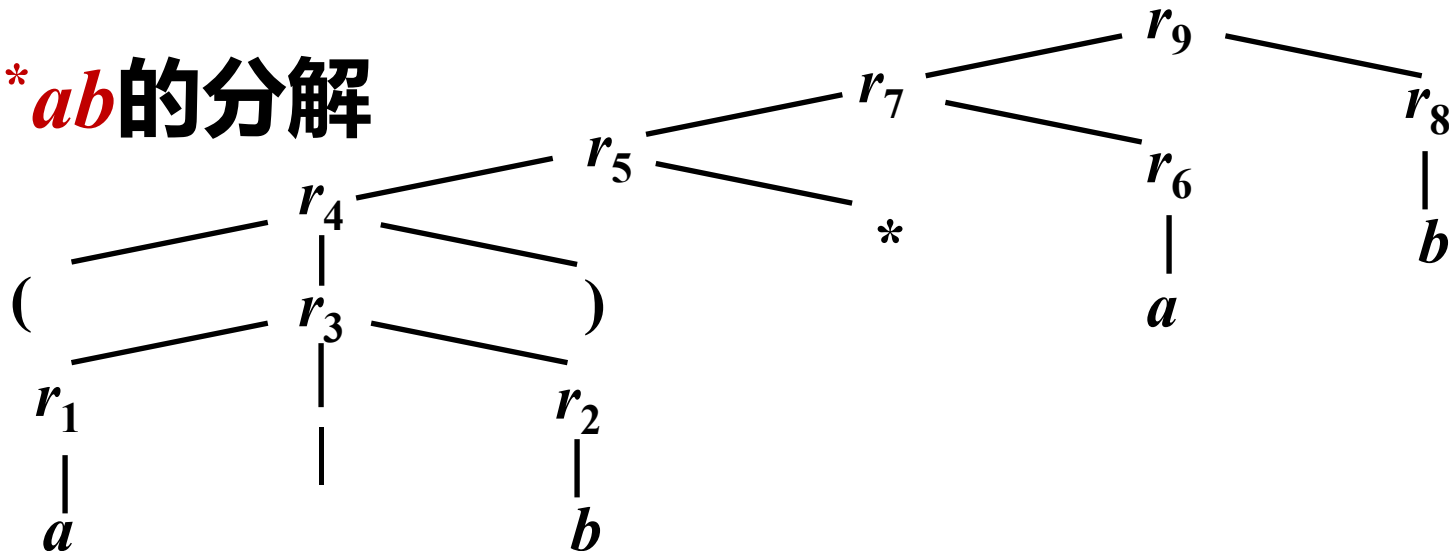
## □ 由本方法产生的NFA具有下列性质:

- ❖  $N(r)$  的每个状态有 (1) 一个其标号为  $\Sigma$  中符号的指向其它状态的转换, 或者 (2) 最多两个指向其它状态的  $\epsilon$  转换





$(a|b)^*ab$  的分解

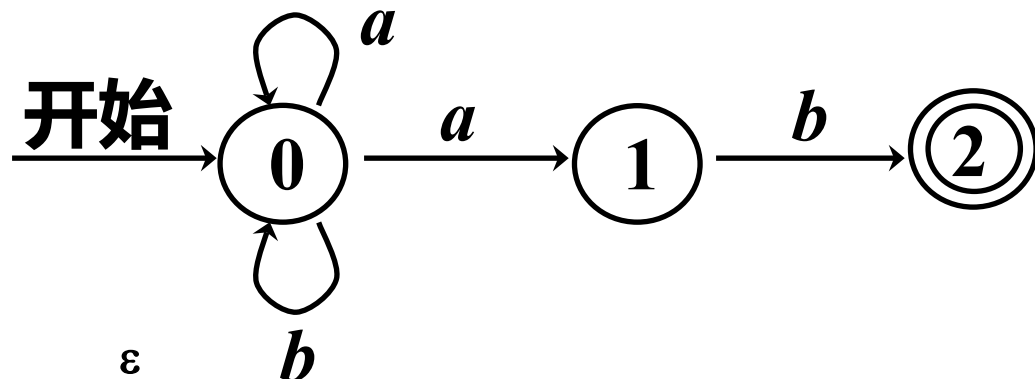




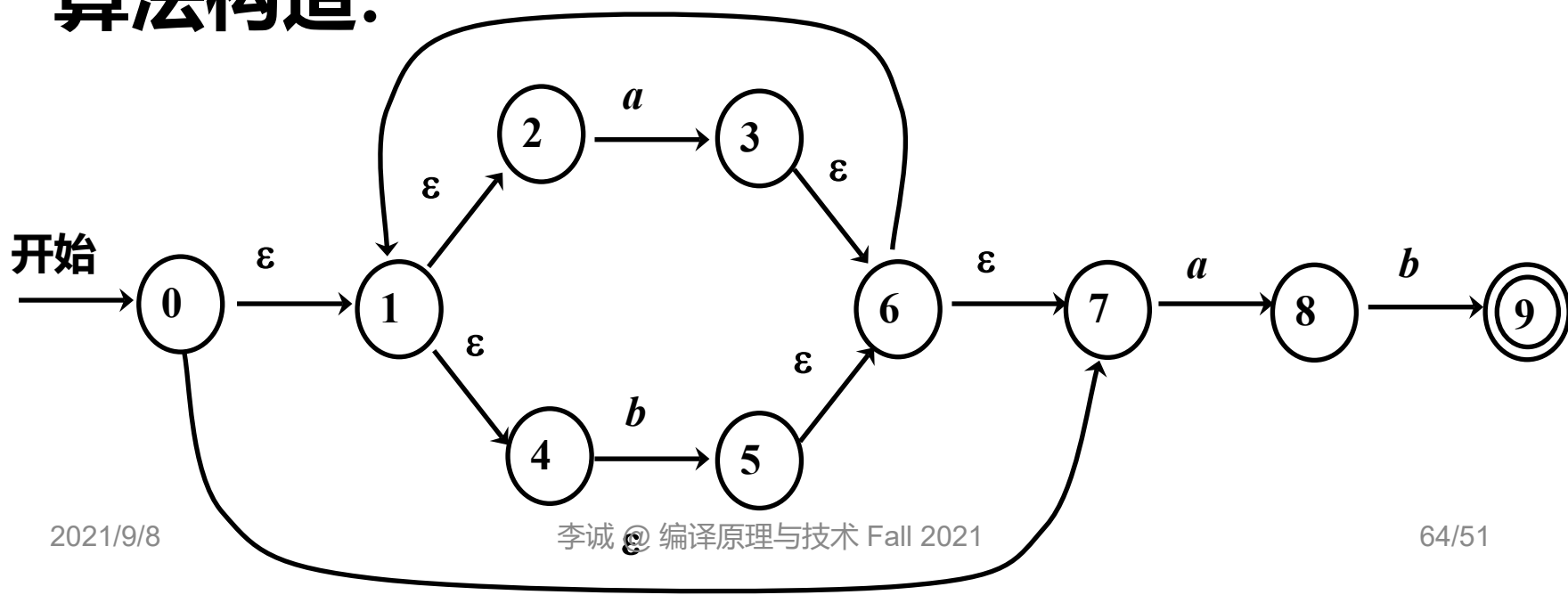


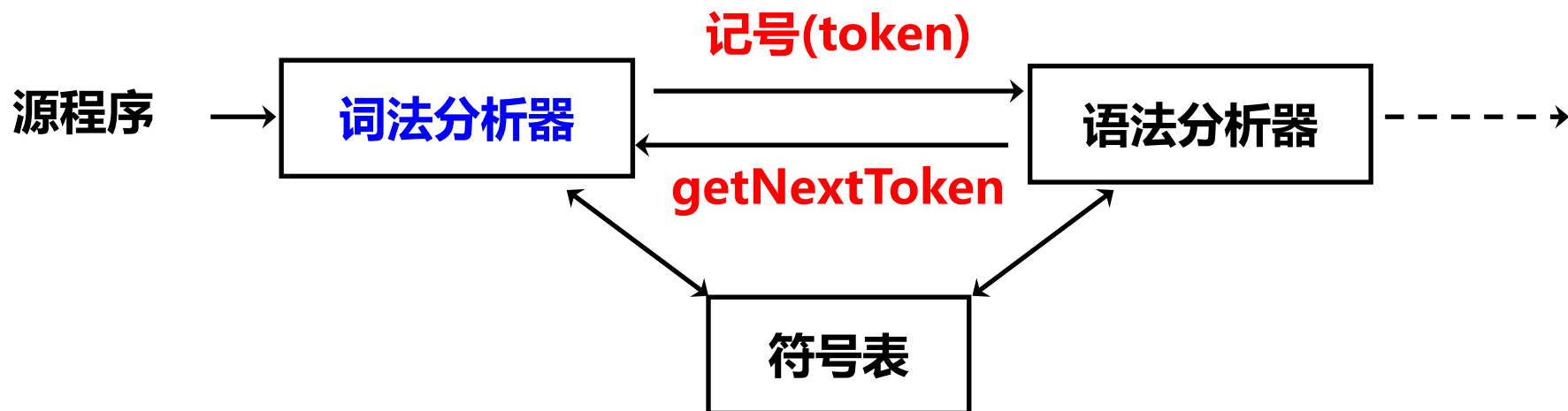
## □ $(a|b)^*ab$ 的两个NFA的比较

手工构造:



算法构造:





## □ 词法分析概述及自动化需要解决的问题

## □ 词法分析器的自动生成

- ❖ 词法单元的描述：正则式
- ❖ 词法单元的识别：转换图
- ❖ 有限自动机：NFA、DFA
- ❖ 正则表达式 → **NFA** → **DFA** → 化简的DFA

## □ NFA到DFA的变换

### □ 子集构造法

- ❖ DFA的一个状态是NFA的一个状态集合
- ❖ 读了输入  $a_1 a_2 \dots a_n$  后,  
NFA能到达的所有状态:  $s_1, s_2, \dots, s_k$ , 则  
DFA到达状态  $\{s_1, s_2, \dots, s_k\}$

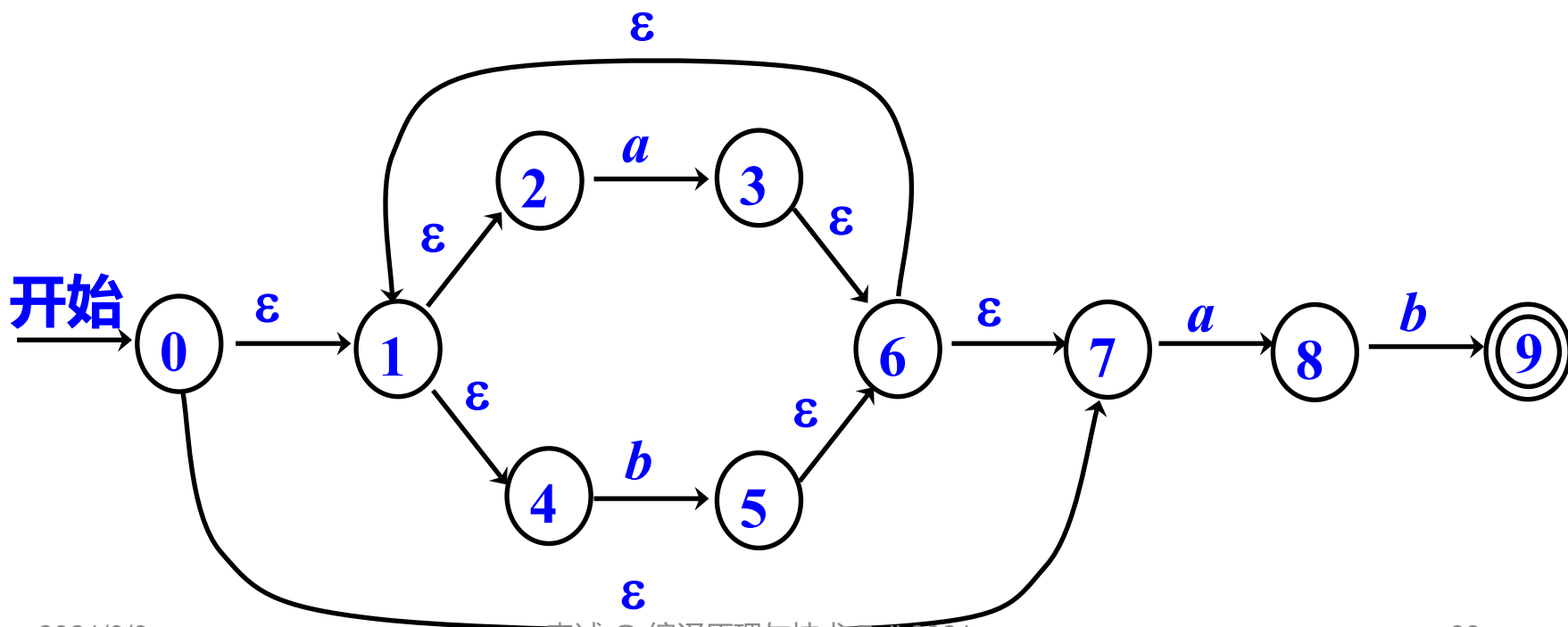


## □ 子集构造法(subset construction)

- ❖  $\epsilon$ -闭包 ( $\epsilon$ -closure) : 状态 $s$ 的 $\epsilon$ -闭包是 $s$ 经 $\epsilon$ 转换所能到达的状态集合
- ❖ NFA的初始状态的 $\epsilon$ -闭包对应于DFA的初始状态
- ❖ 针对每个DFA状态 – NFA状态子集 $A$ , 求输入每个 $a_i$ 后能到达的NFA状态的 $\epsilon$ -闭包并集 ( $\epsilon$ -closure(move( $A, a_i$ ))), 该集合对应于DFA中的一个已有状态, 或者是一个要新加的DFA状态

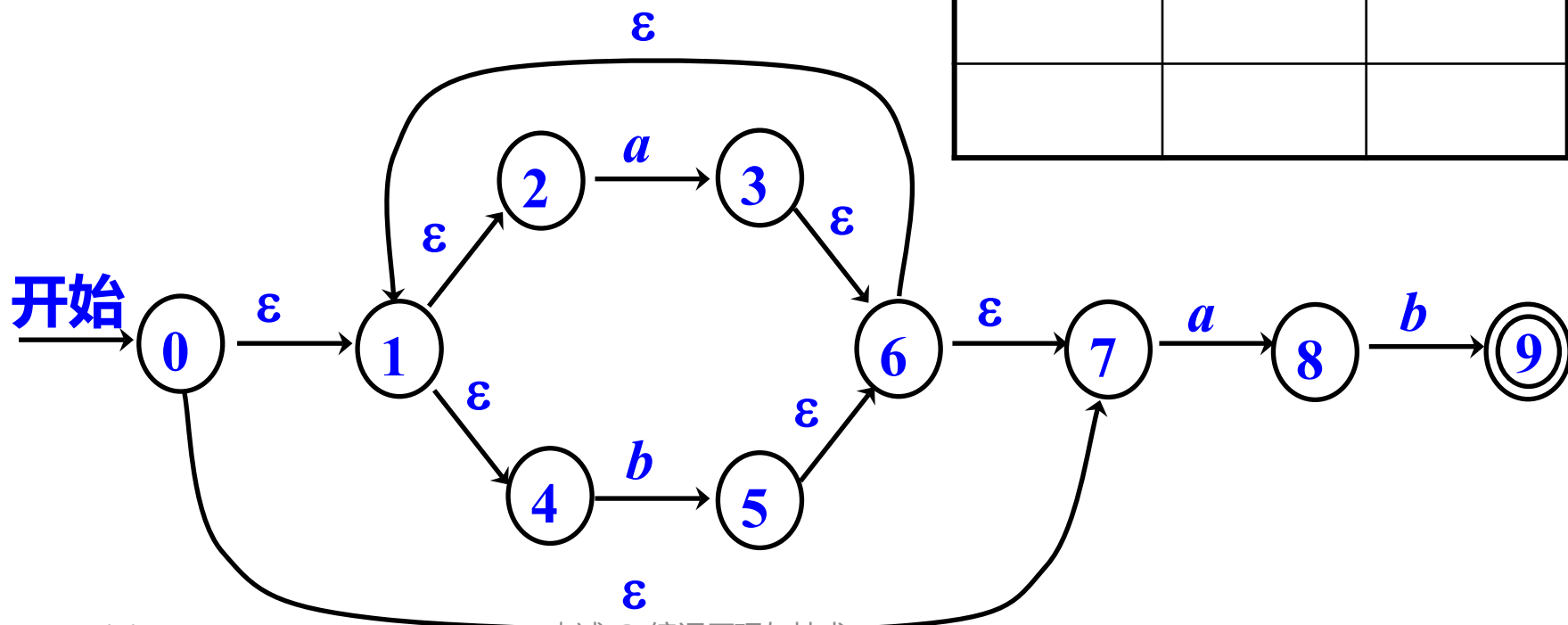


□ 例  $(a|b)^*ab$ , NFA如下, 把它变换为DFA





# NFA到DFA的变换



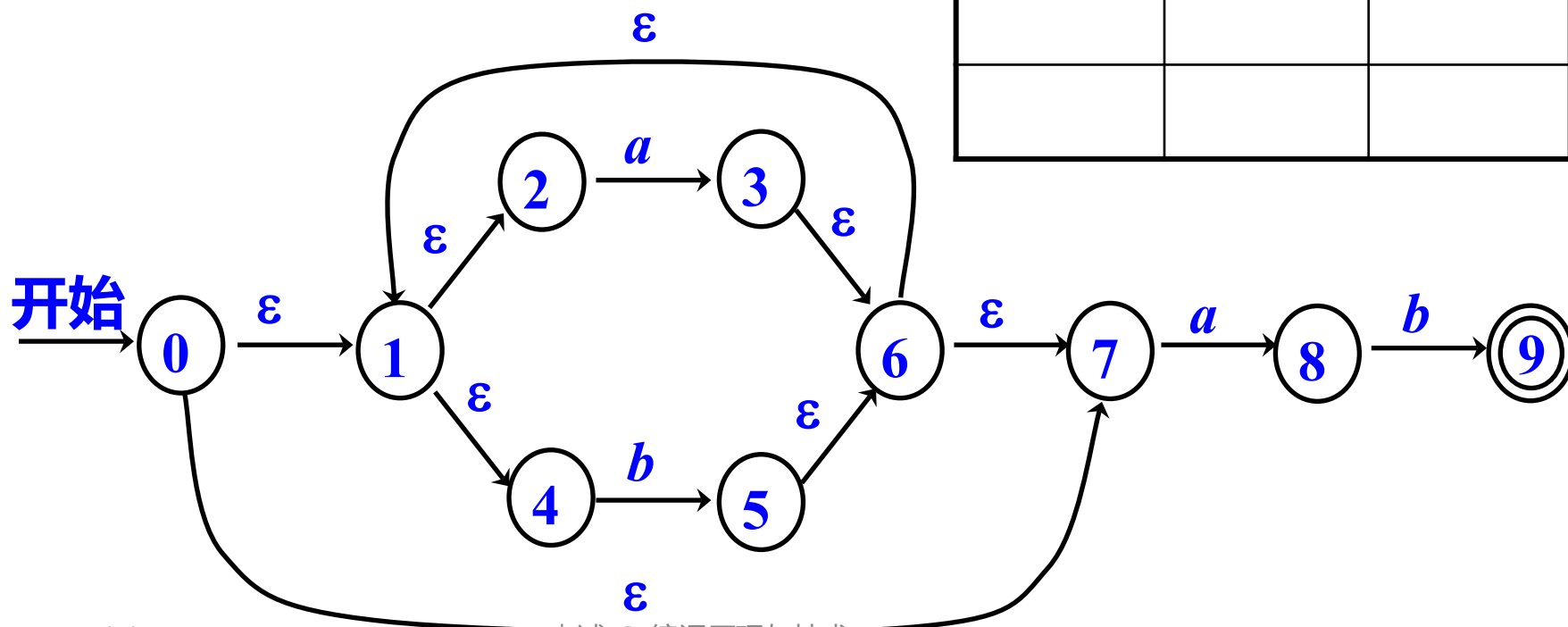


# NFA到DFA的变换



$$A = \{0, 1, 2, 4, 7\}$$

状态	输入符号	
	<i>a</i>	<i>b</i>
<i>A</i>		





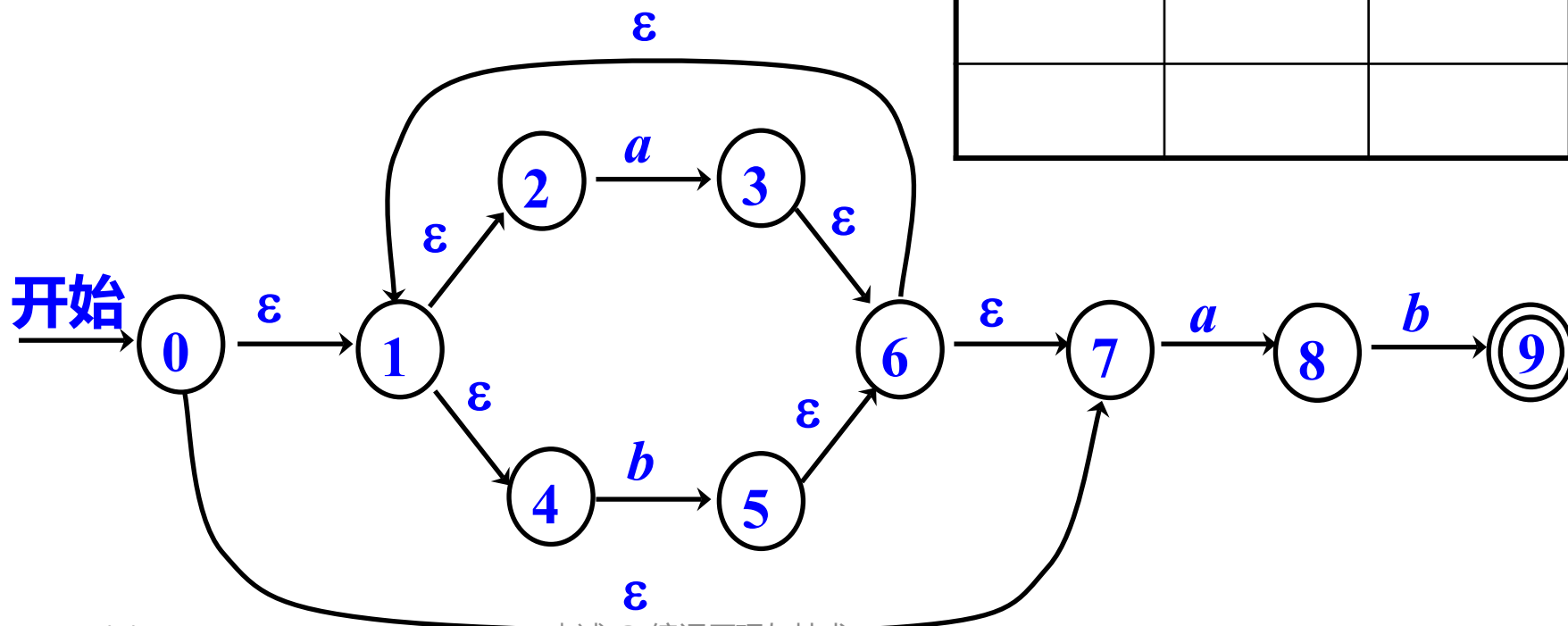
# NFA到DFA的变换



$A = \{0, 1, 2, 4, 7\}$

$B = \{1, 2, 3, 4, 6, 7, 8\}$

状态	输入符号	
	$a$	$b$
$A$	$B$	







# NFA到DFA的变换

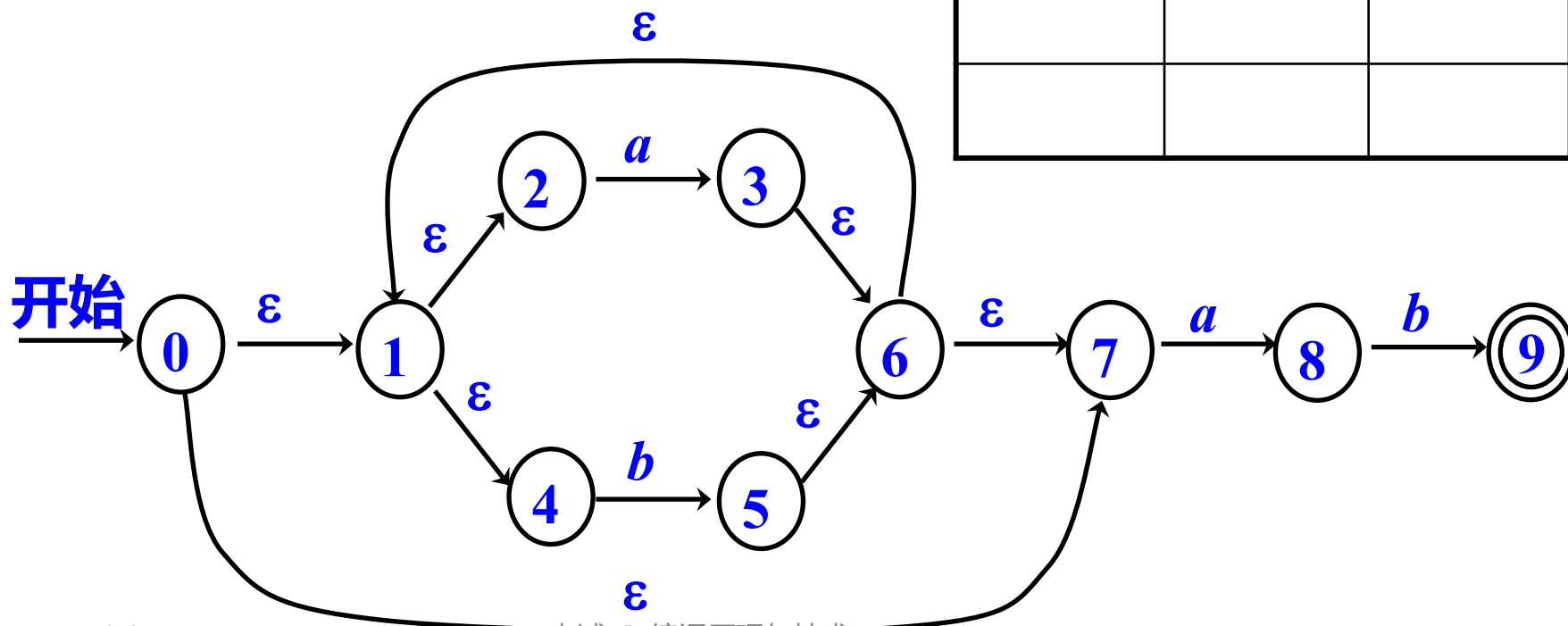


$A = \{0, 1, 2, 4, 7\}$

$B = \{1, 2, 3, 4, 6, 7, 8\}$

$C = \{1, 2, 4, 5, 6, 7\}$

状态	输入符号	
	$a$	$b$
$A$	$B$	$C$





# NFA到DFA的变换

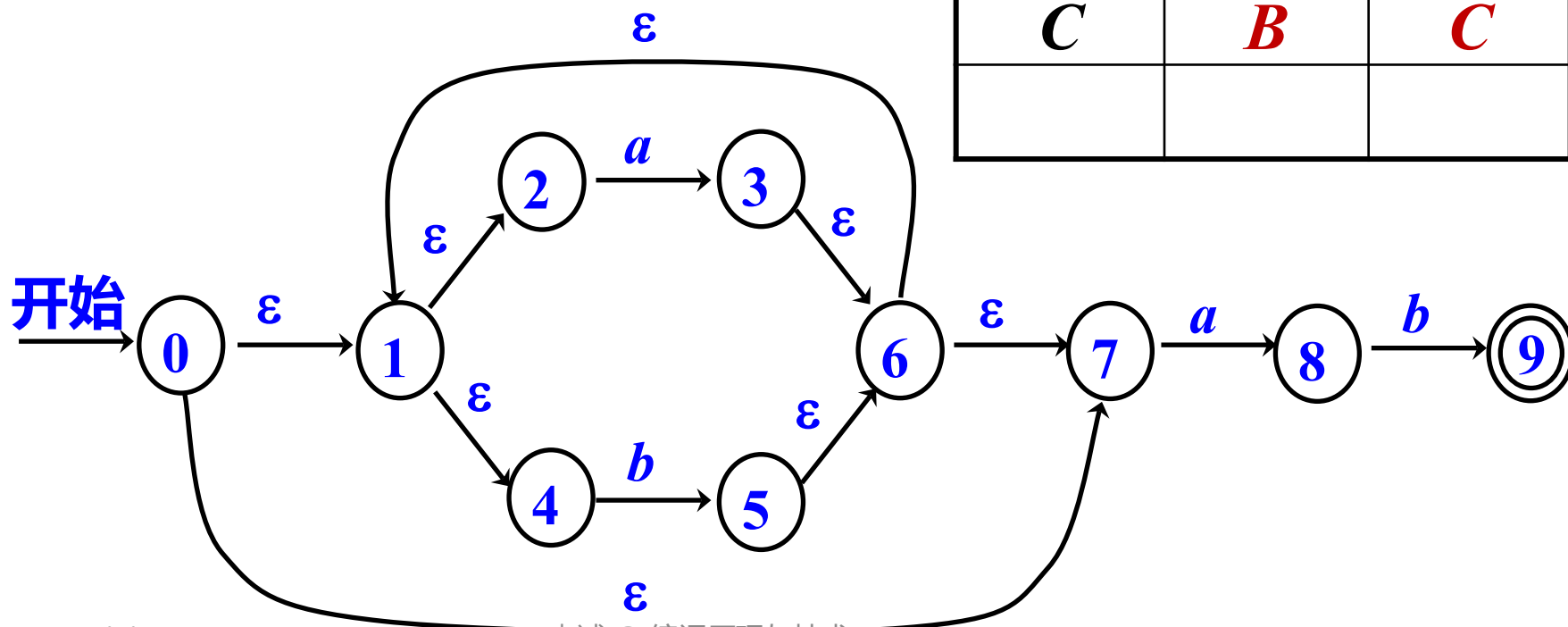


$$A = \{0, 1, 2, 4, 7\}$$

$$B = \{1, 2, 3, 4, 6, 7, 8\}$$

$$C = \{1, 2, 4, 5, 6, 7\}$$

状态	输入符号	
	<i>a</i>	<i>b</i>
<i>A</i>	<i>B</i>	<i>C</i>
<i>B</i>	<i>B</i>	
<i>C</i>	<i>B</i>	<i>C</i>





# NFA到DFA的变换



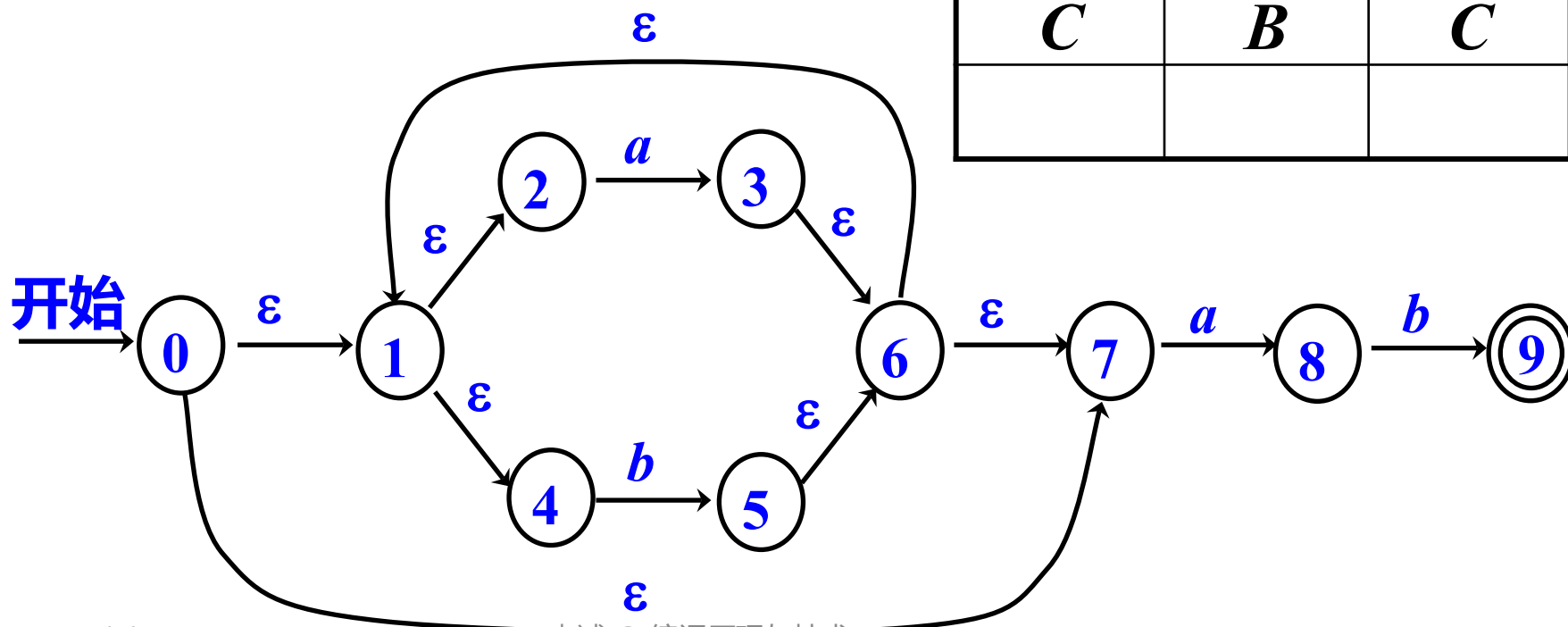
$$A = \{0, 1, 2, 4, 7\}$$

$$B = \{1, 2, 3, 4, 6, 7, 8\}$$

$$C = \{1, 2, 4, 5, 6, 7\}$$

$$D = \{1, 2, 4, 5, 6, 7, 9\}$$

状态	输入符号	
	<i>a</i>	<i>b</i>
<i>A</i>	<i>B</i>	<i>C</i>
<i>B</i>	<i>B</i>	
<i>C</i>	<i>B</i>	<i>C</i>





# NFA到DFA的变换



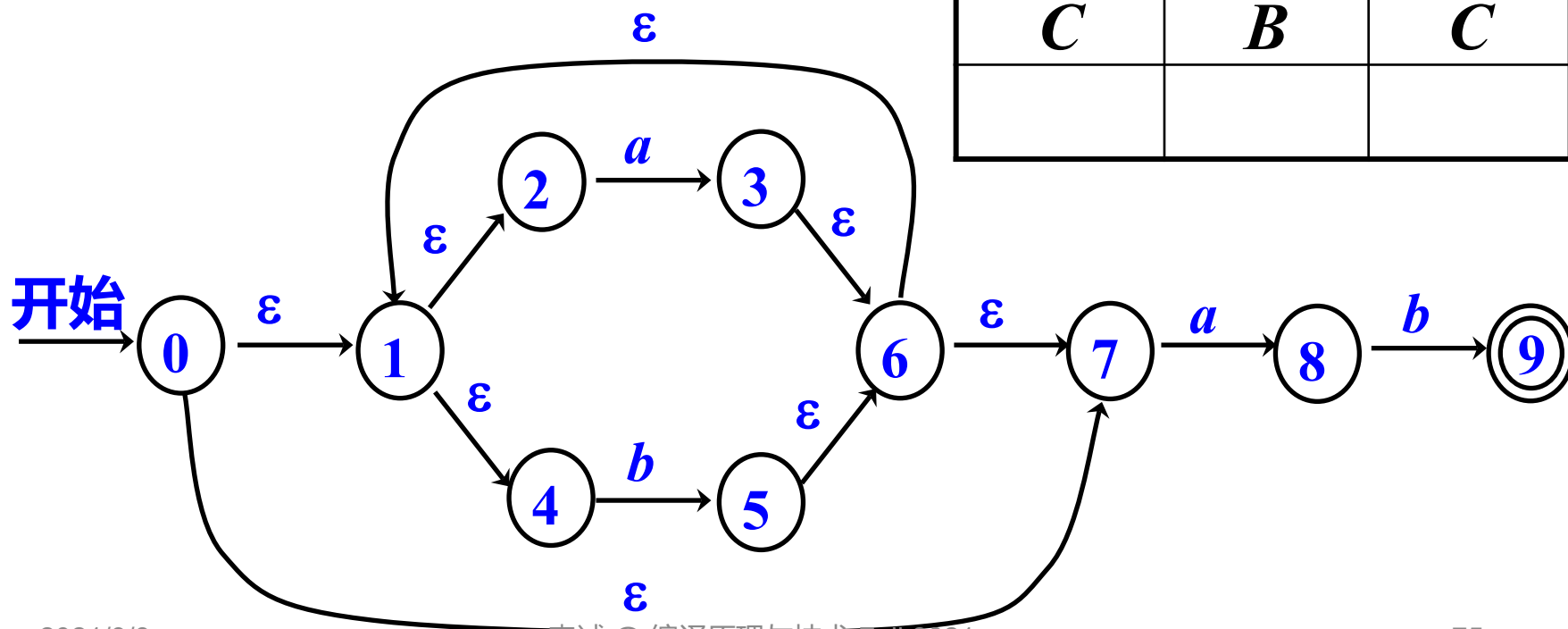
$$A = \{0, 1, 2, 4, 7\}$$

$$B = \{1, 2, 3, 4, 6, 7, 8\}$$

$$C = \{1, 2, 4, 5, 6, 7\}$$

$$D = \{1, 2, 4, 5, 6, 7, 9\}$$

状态	输入符号	
	<i>a</i>	<i>b</i>
<i>A</i>	<i>B</i>	<i>C</i>
<i>B</i>	<i>B</i>	<i>D</i>
<i>C</i>	<i>B</i>	<i>C</i>





# NFA到DFA的变换



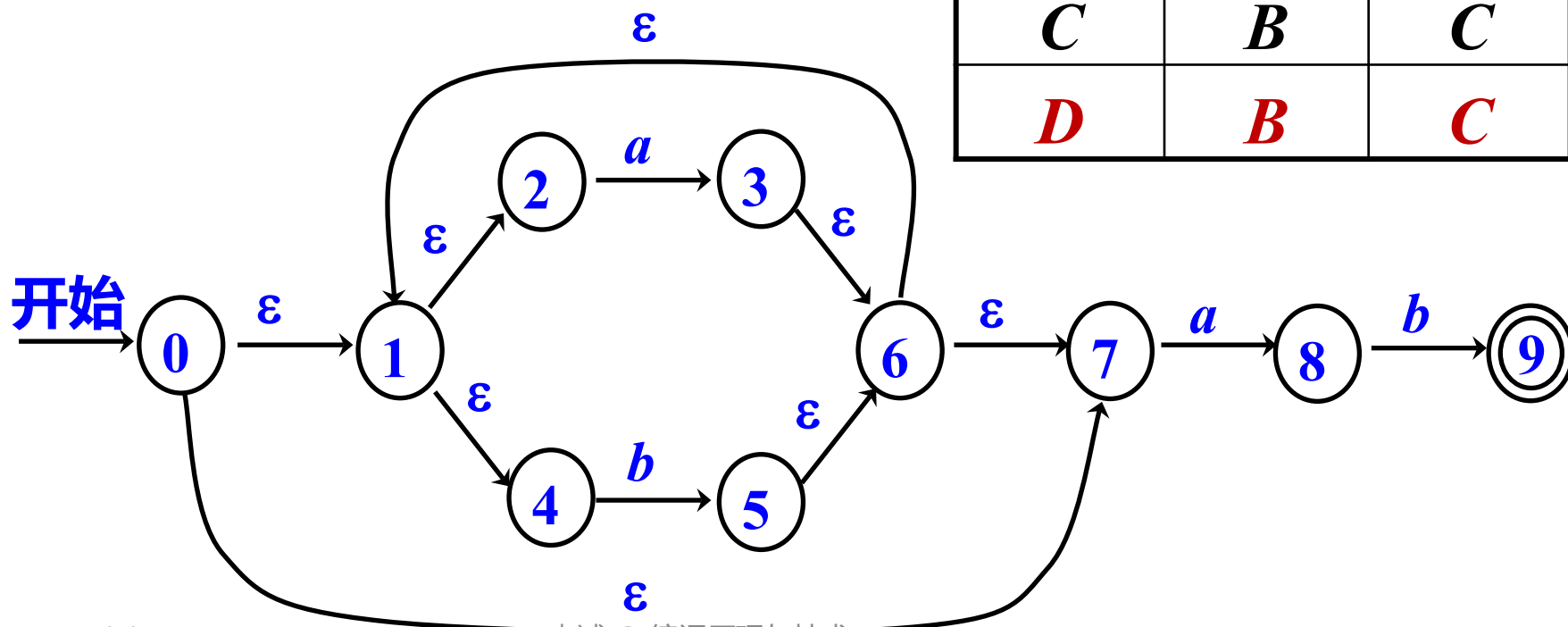
$$A = \{0, 1, 2, 4, 7\}$$

$$B = \{1, 2, 3, 4, 6, 7, 8\}$$

$$C = \{1, 2, 4, 5, 6, 7\}$$

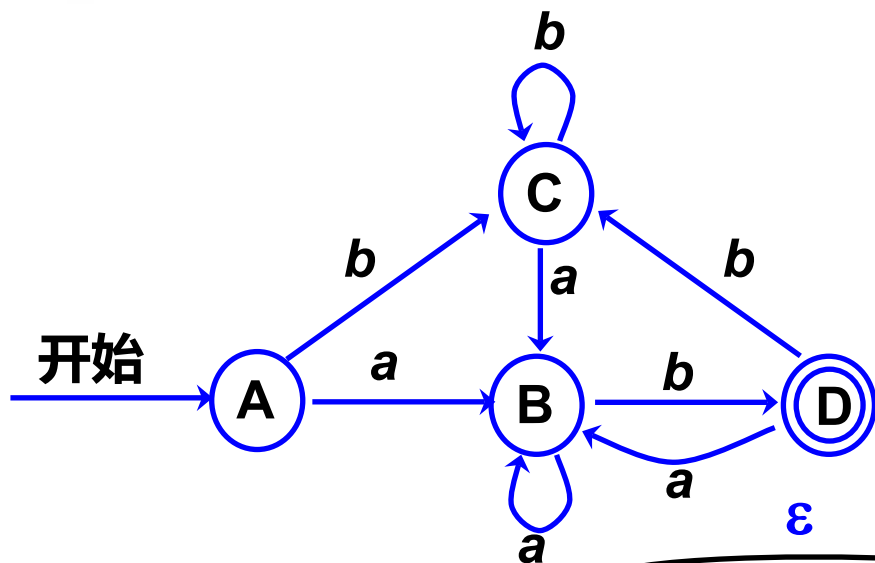
$$D = \{1, 2, 4, 5, 6, 7, 9\}$$

状态	输入符号	
	<i>a</i>	<i>b</i>
<i>A</i>	<i>B</i>	<i>C</i>
<i>B</i>	<i>B</i>	<i>D</i>
<i>C</i>	<i>B</i>	<i>C</i>
<i>D</i>	<i>B</i>	<i>C</i>

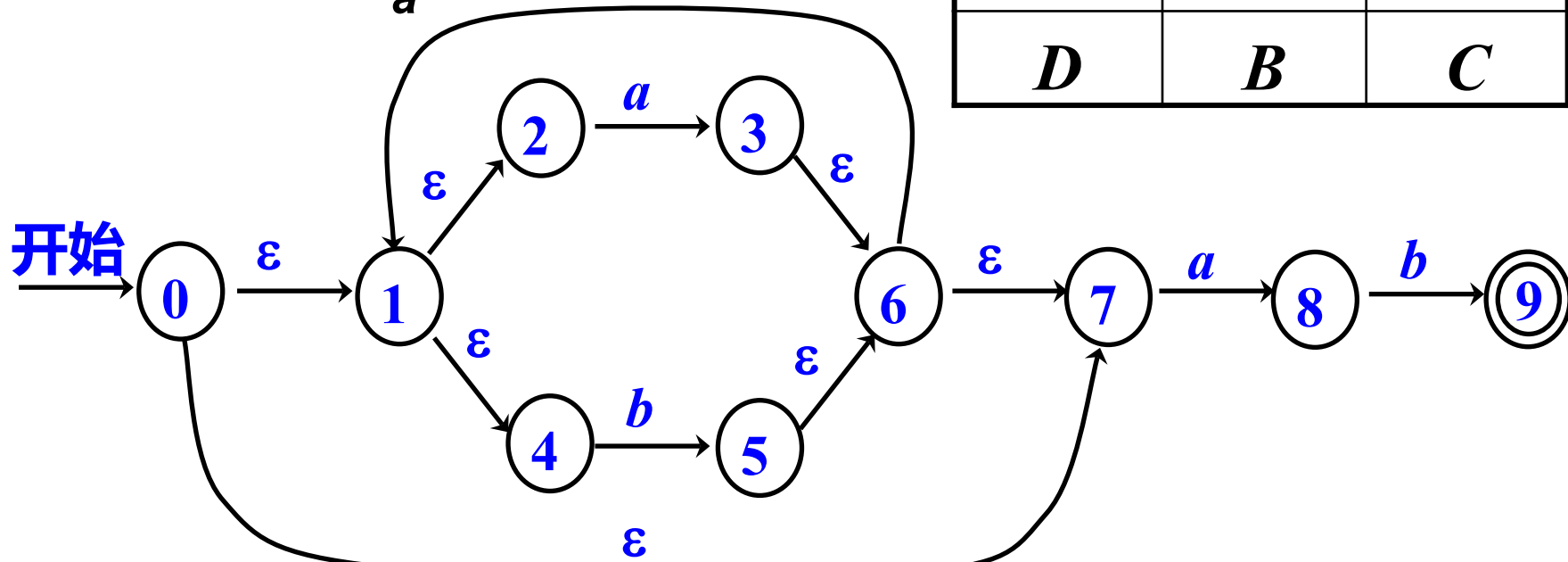




# NFA到DFA的变换



状态	输入符号	
	<i>a</i>	<i>b</i>
<i>A</i>	<i>B</i>	<i>C</i>
<i>B</i>	<i>B</i>	<i>D</i>
<i>C</i>	<i>B</i>	<i>C</i>
<i>D</i>	<i>B</i>	<i>C</i>

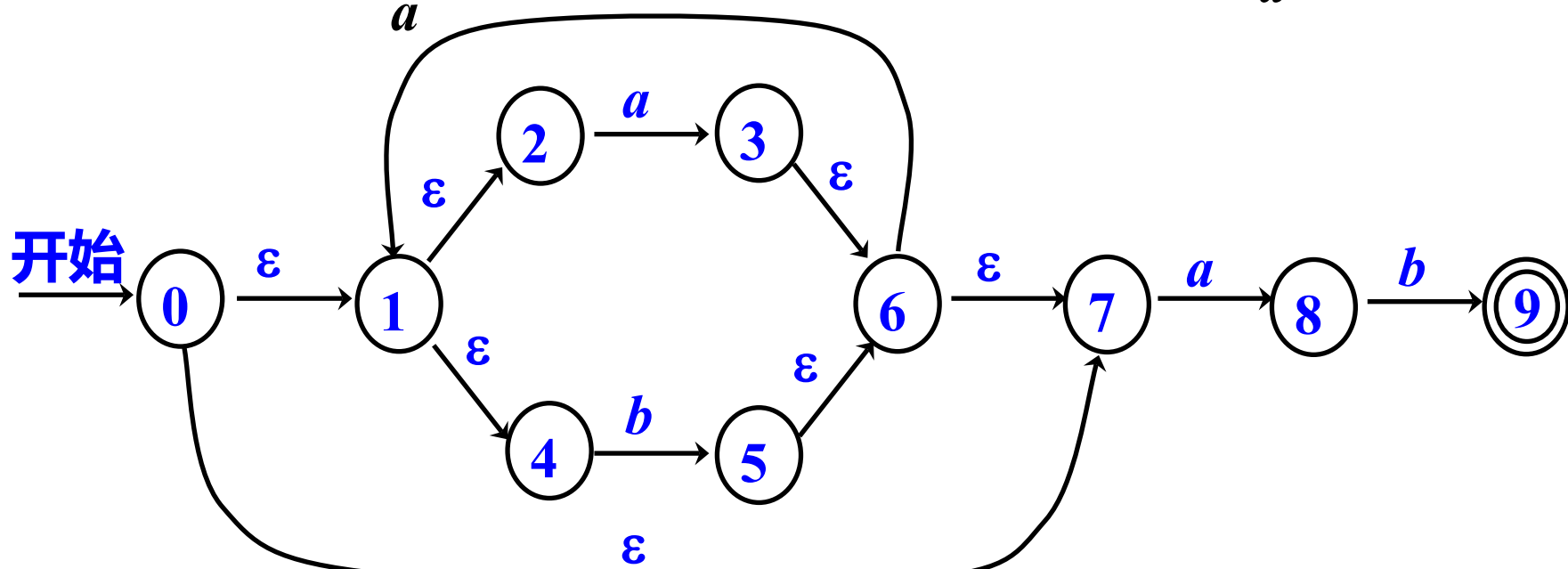
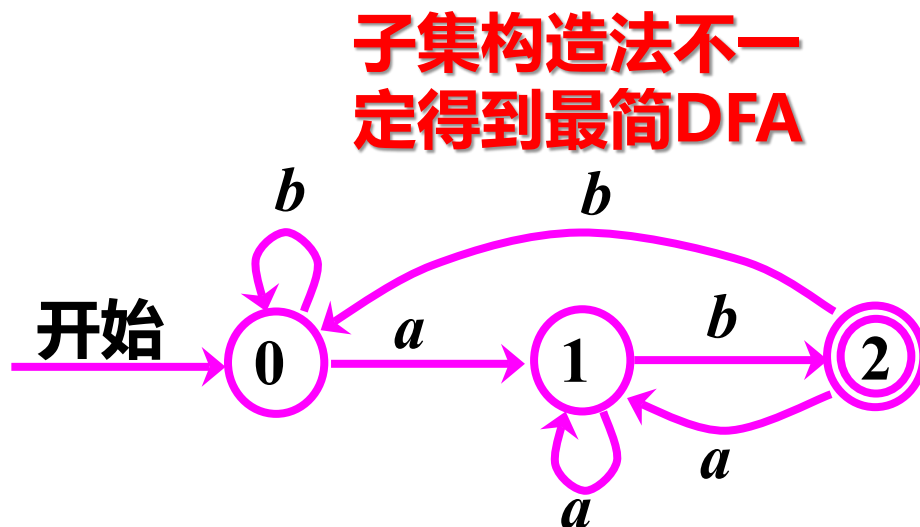
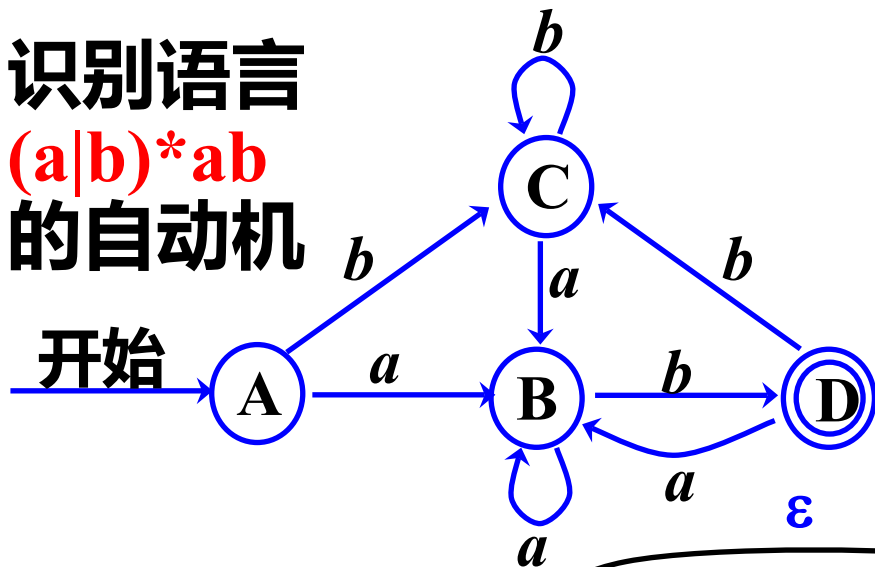


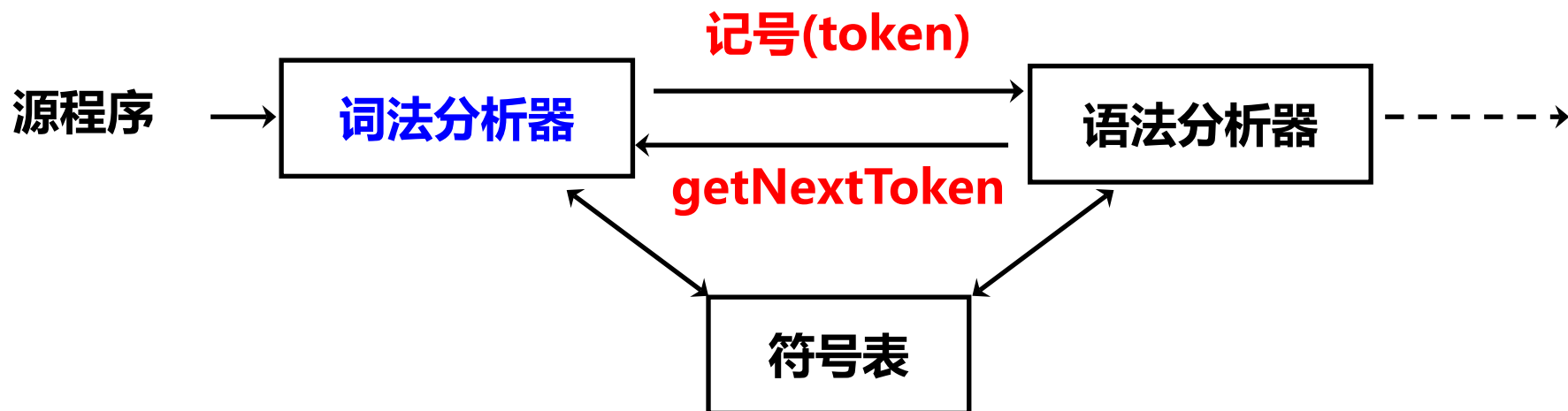


# NFA到DFA的变换



识别语言  
 $(a|b)^*ab$   
的自动机





## □ 词法分析概述及自动化需要解决的问题

## □ 词法分析器的自动生成

- ❖ 词法单元的描述：正则式
- ❖ 词法单元的识别：转换图
- ❖ 有限自动机：NFA、DFA
- ❖ 正则表达式 → NFA → **DFA** → 化简的**DFA**





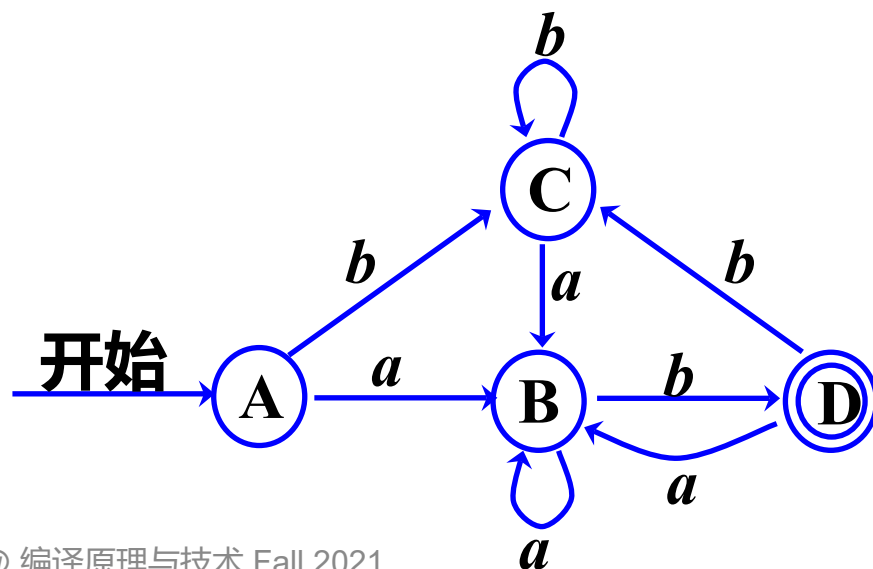
## □ A和B是可区别的状态

- ❖ 从A出发，读过单字符b构成的串，到达非接受状态C，而从B出发，读过串b，到达接受状态D

## □ A和C是不可区别的状态

- ❖ 无任何串可用来像上面这样区别它们

可区别的状态要  
分开对待



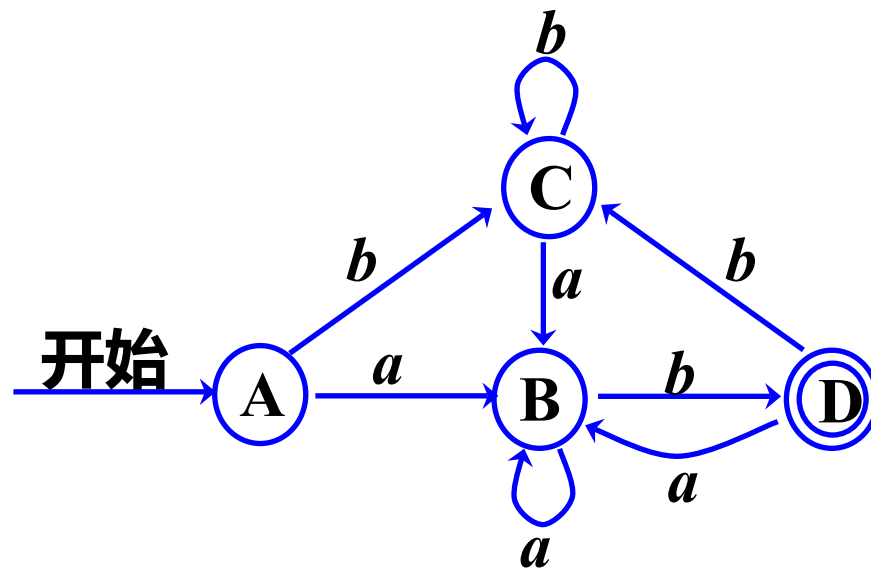


## 1. 按是否是接受状态来区分

$\{A, B, C\}, \{D\}$

$\text{move}(\{A, B, C\}, a) = \{B\}$

$\text{move}(\{A, B, C\}, b) = \{C, D\}$





## 1. 按是否是接受状态来区分

$\{A, B, C\}, \{D\}$

$\text{move}(\{A, B, C\}, a) = \{B\}$

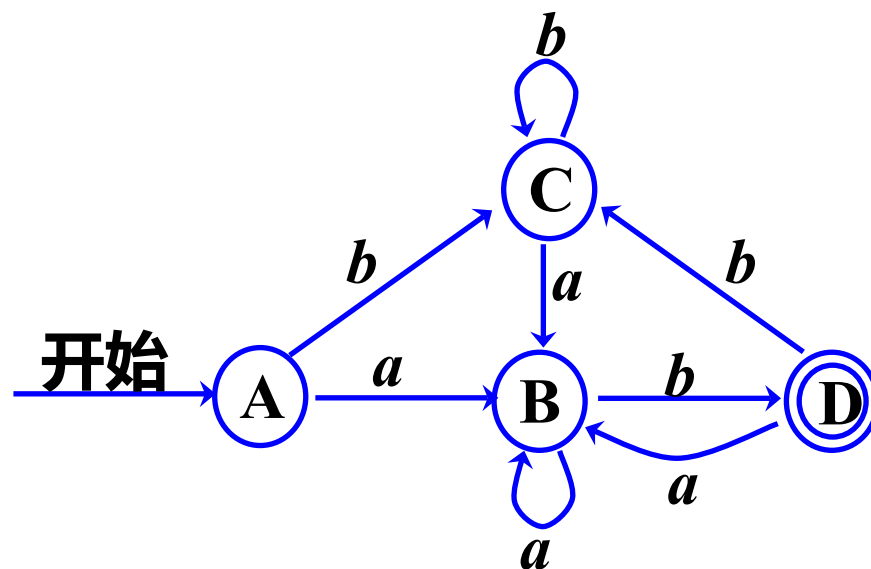
$\text{move}(\{A, B, C\}, b) = \{C, D\}$

## 2. 继续分解

$\{A, C\}, \{B\}, \{D\}$

$\text{move}(\{A, C\}, a) = \{B\}$

$\text{move}(\{A, C\}, b) = \{C\}$



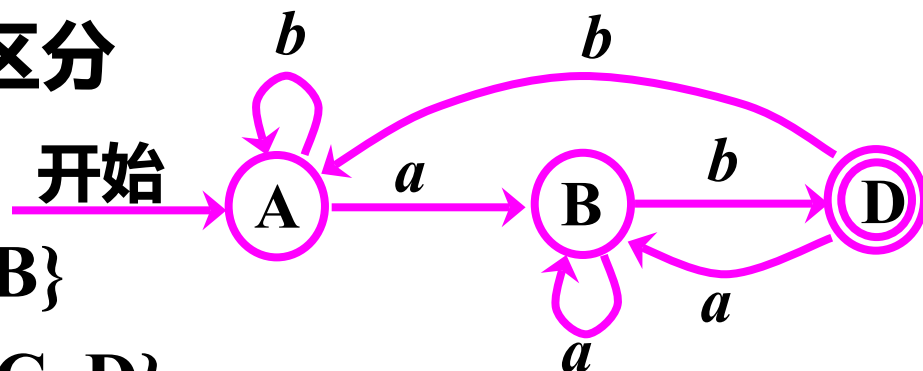


## 1. 按是否是接受状态来区分

$\{A, B, C\}, \{D\}$

$\text{move}(\{A, B, C\}, a) = \{B\}$

$\text{move}(\{A, B, C\}, b) = \{C, D\}$

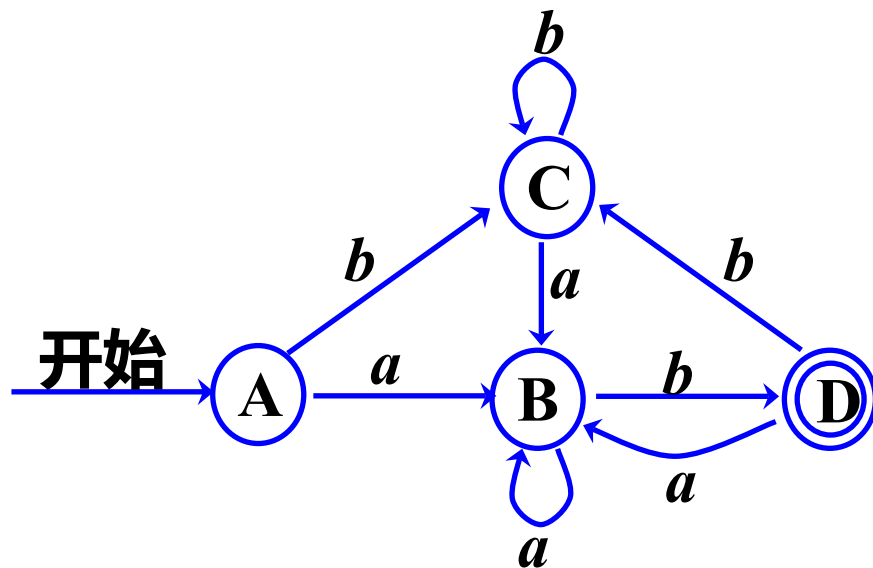


## 2. 继续分解

$\{A, C\}, \{B\}, \{D\}$

$\text{move}(\{A, C\}, a) = \{B\}$

$\text{move}(\{A, C\}, b) = \{C\}$





□ 正则表达式  $(a|b)^*$  与  $(a^*|b^*)^*$  是否等价？

❖ 提示：可利用其最简化DFA的