编译原理

正规式与有限自动机的等价性

编译原理

正规式与有限自动机的等价性

——回顾

词法分析

- 词法分析器的设计
- 正规表达式与有限自动机
- 词法分析器的自动产生--LEX

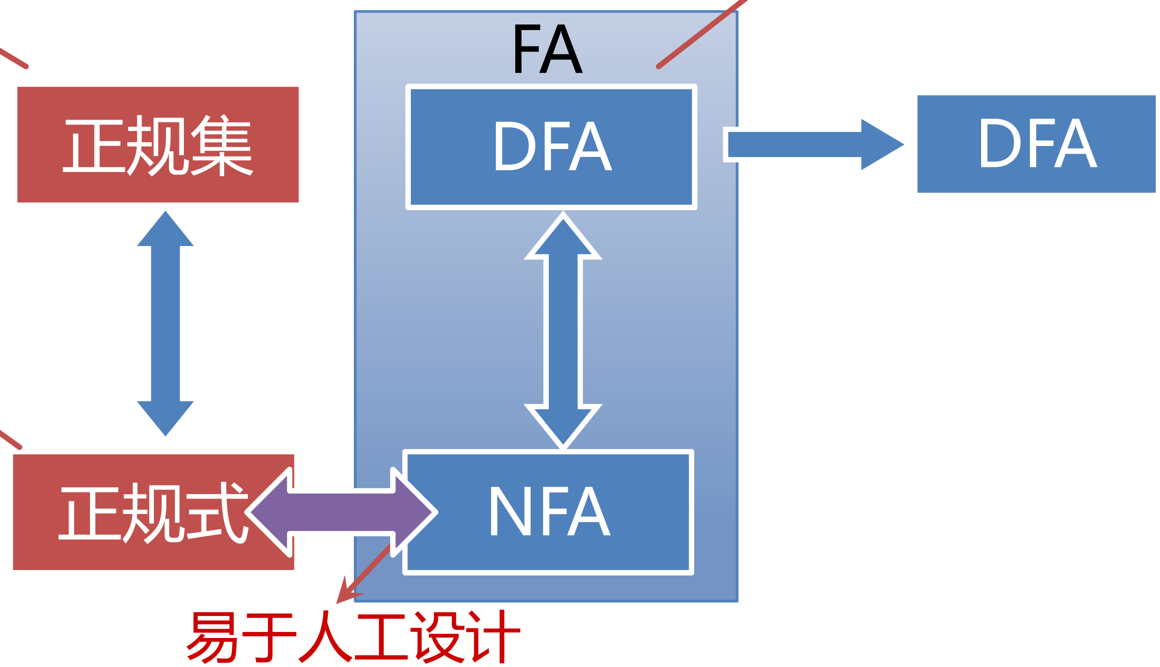
单词符号	种别编码	助忆符	内码值
DIM	1	$DIM	-
IF	2	$IF	-

正规式、正

```
curState = 初态
GetChar();
while( stateTrans[curState][ch]有定义){
    //存在后继状态，读入、拼接
    Concat();
    //转换入下一状态，读入下一字符
    curState= stateTrans[curState][ch];
    if curState是终态 then 返回strToken中的单词
    GetChar( );
}
```

DIM,IF, DO,STOP,END
number, name, age
125, 2169
…

DIM
IF
DO
STOP
END
letter(letter|digit)*
digit(digit)*



正规式与有限自动机的等价性

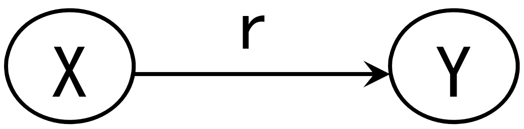
- 一个正规式r与一个有限自动机M等价
  - L(r)=L(M)
- FA ->正规式
  - 对任何FA M，都存在一个正规式r，使得L(r)=L(M)。
- 正规式 -> FA
  - 对任何正规式r，都存在一个FA M，使得L(M)=L(r)。

编译原理

正规式与有限自动机的等价性
——为NFA构造正规式

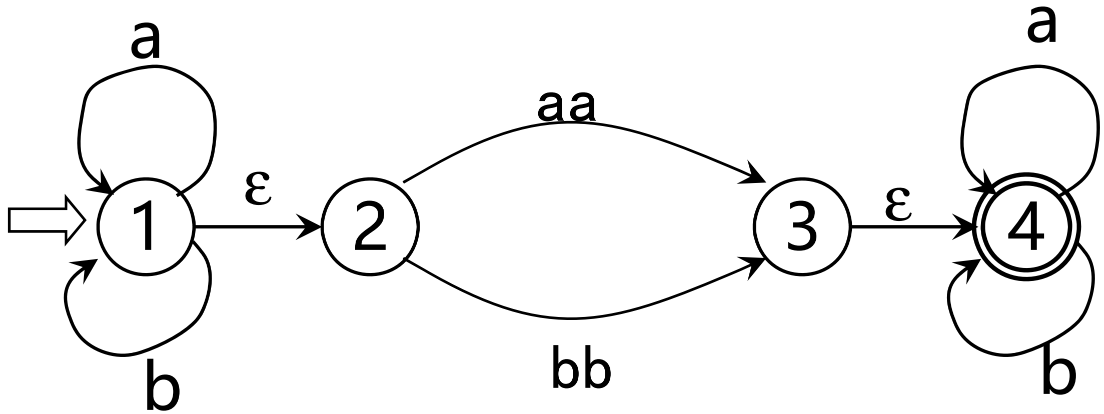
为NFA构造正规式

- 对转换图概念拓广，令每条弧可用一个正规式作标记。
- 证明： 对Σ上任一NFA M，都存在一个Σ上的正规式r，使得L(r)=L(M)。



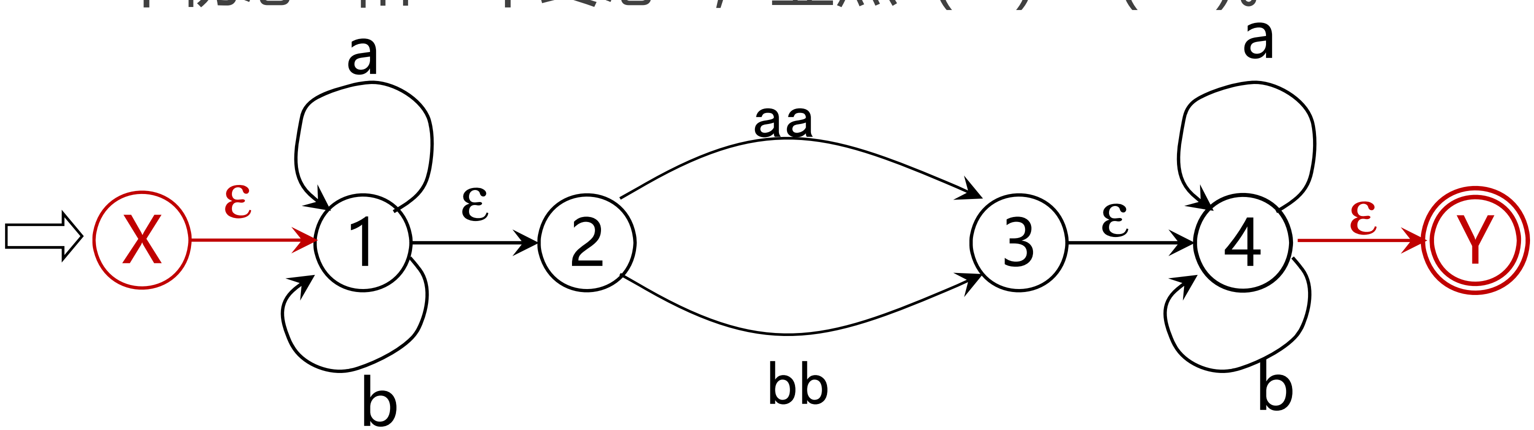
为NFA构造正规式

- 假定NFA $M=<S, \Sigma, \delta, S_0, F>$，我们对M的状态转换图进行以下改造：
  - 在M的转换图上加进两个状态X和Y，从X用ε弧连接到M的所有初态结点，从M的所有终态结点用ε弧连接到Y，从而形成一个新的NFA，记为M’，它只有一个初态X和一个终态Y，显然L(M)=L(M’)。



为NFA构造正规式

- 假定NFA M=<S, Σ, δ, S_0, F>，我们对M的状态转换图进行以下改造：
  - 在M的转换图上加进两个状态X和Y，从X用ε弧连接到M的所有初态结点，从M的所有终态结点用ε弧连接到Y，从而形成一个新的NFA，记为M'，它只有一个初态X和一个终态Y，显然L(M)=L(M')。

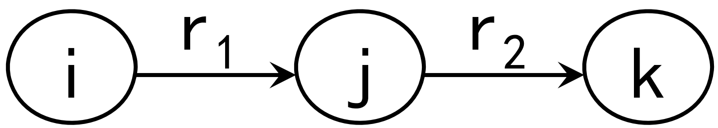


M'

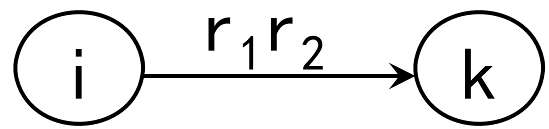
为NFA构造正规式

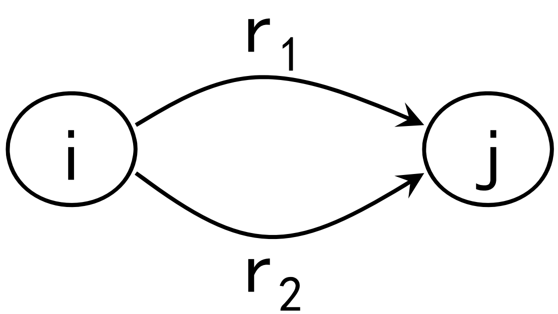
- 假定NFA M=<S, Σ, δ, S_0, F>，我们对M的状态转换图进行以下改造：
  - 在M的转换图上加进两个状态X和Y，从X用ε弧连接到M的所有初态结点，从M的所有终态结点用ε弧连接到Y，从而形成一个新的NFA，记为M'，它只有一个初态X和一个终态Y，显然L(M)=L(M')。
  - 然后，反复使用下面的三条规则，逐步消去结点，直到只剩下X和Y为止。

为NFA构造正规式

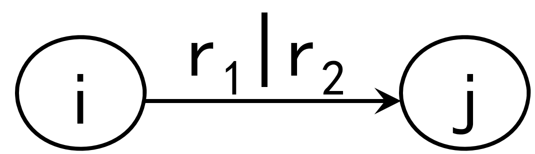


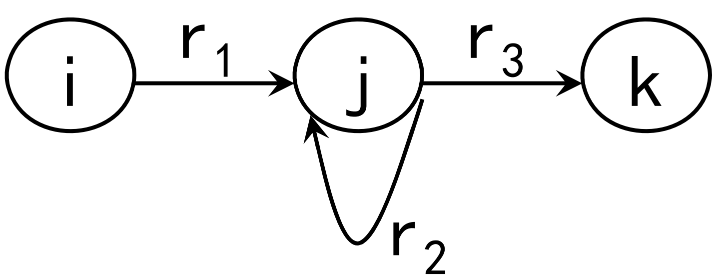
代之为



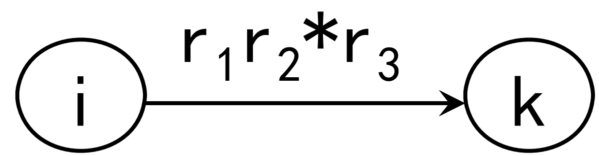


代之为

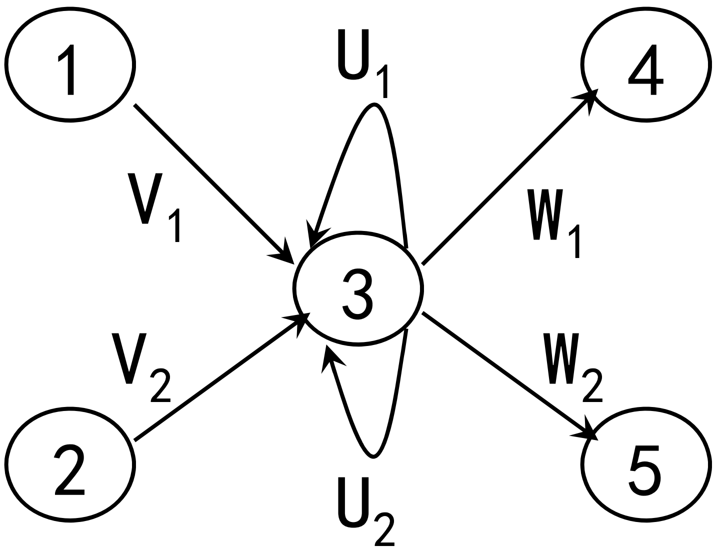


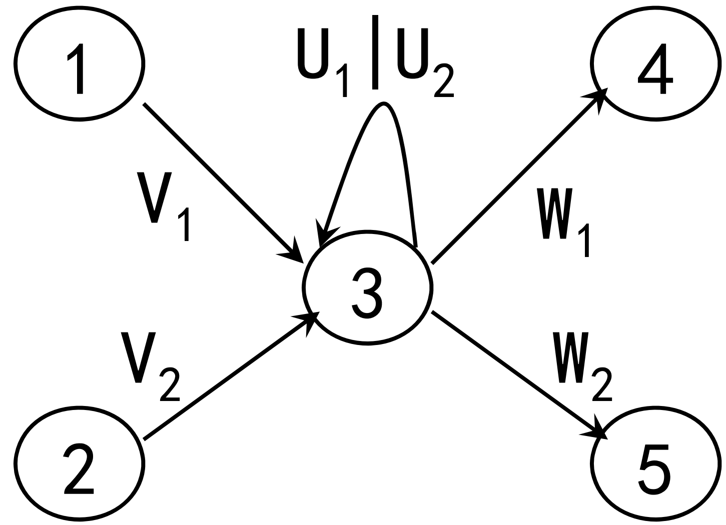


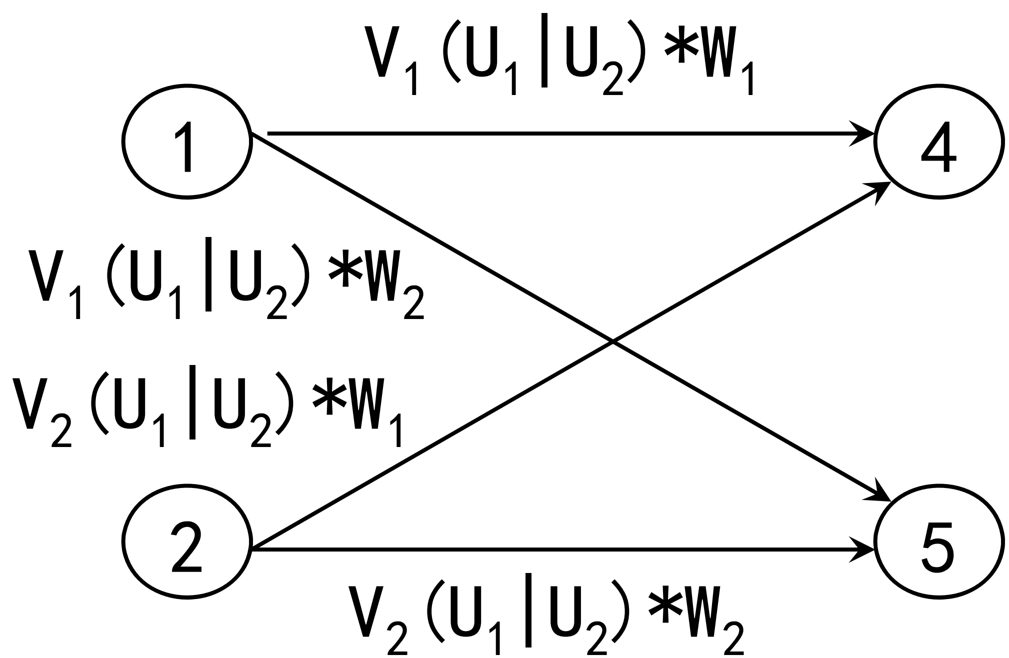
代之为



为NFA构造正规式

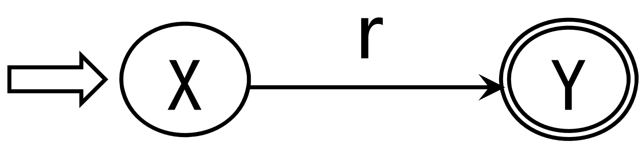

1
2
3
4
5
U_1
U_2
V_1
V_2
W_1
W_2


1
2
3
4
5
$U_1|U_2$
V_1
V_2
W_1
W_2


1
2
4
5
$V_1(U_1|U_2)*W_1$
$V_1(U_1|U_2)*W_2$
$V_2(U_1|U_2)*W_1$
$V_2(U_1|U_2)*W_2$

为NFA构造正规式

- 最后，X到Y的弧上标记的正规式即为所构造的正规式r



- 显然L(r)=L(M’)=L(M)
- 得证： 对Σ上任一NFA M，都存在一个Σ上的正规式r，使得L(r)=L(M)。

1. 对任何FA M，都存在一个正规式r，使得L(r)=L(M)。
2. 对任何正规式r，都存在一个FA M，使得L(M)=L(r)。

编译原理

正规式与有限自动机的等价性
——为正规式构造NFA

为正规式构造NFA

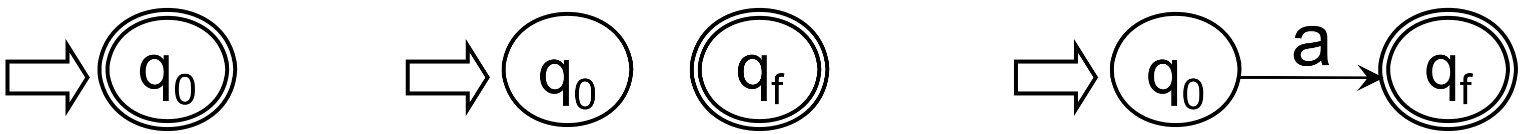
- 定理：对任何正规式r，都存在一个FA M，使得L(M)=L(r)。
- 定理: 对于Σ上的正规式r，都存在一个NFA M，使L(M)=L(r)，并且M只有一个初态和一个终态，而且没有从终态出发的箭弧。

为正规式构造NFA

- 对给定正规式r中的运算符数目进行归纳
  - 验证r中的运算符数目为0时，结论成立。
  - 假设结论对于运算符数目少于k(k≥1)的正规式成立
  - 基于该假设，证明结论对于运算符数目为k的正规式成立。

为正规式构造NFA

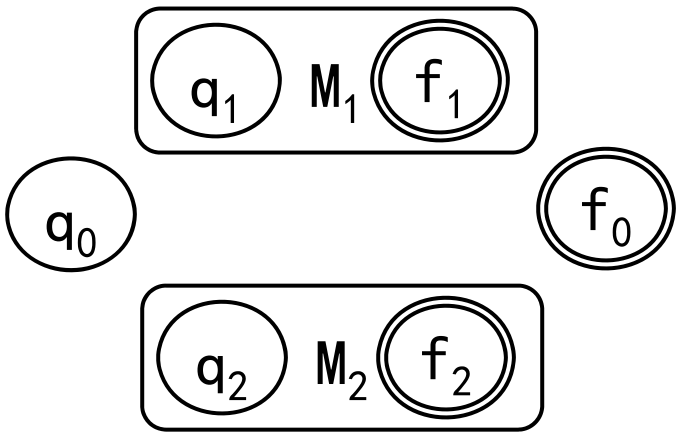
- 若r具有零个运算符，则r=ε或r=φ或r=a，其中a∈Σ。
- 针对上述3类正规式r，分别按照下图构造NFA M，M只有一个初态和一个终态，而且没有从终态出发的箭弧，而且使L(M)和对应的L(r)相等。



为正规式构造NFA

- $r=r_1|r_2$
- $r=r_1.r_2$
- $r=r_1{}^*$

- 假设对于运算符数目少于k(k≥1)的正规式成立。
- 当r中含有k个运算符时，r有三种情形:
  - 情形1：$r=r_1|r_2$，r_1和r_2中运算符数目少于k。从而，由归纳假设，对r_i存在$M_i=<S_i, \Sigma_i, \delta_i, q_i, \{f_i\}>$，使得$L(M_i)=L(r_i)$，并且$M_i$没有从终态出发的箭弧（i=1,2）。不妨设$S_1 \cap S_2=\phi$，在$S_1 \cup S_2$中加入两个新状态$q_0$，$f_0$。



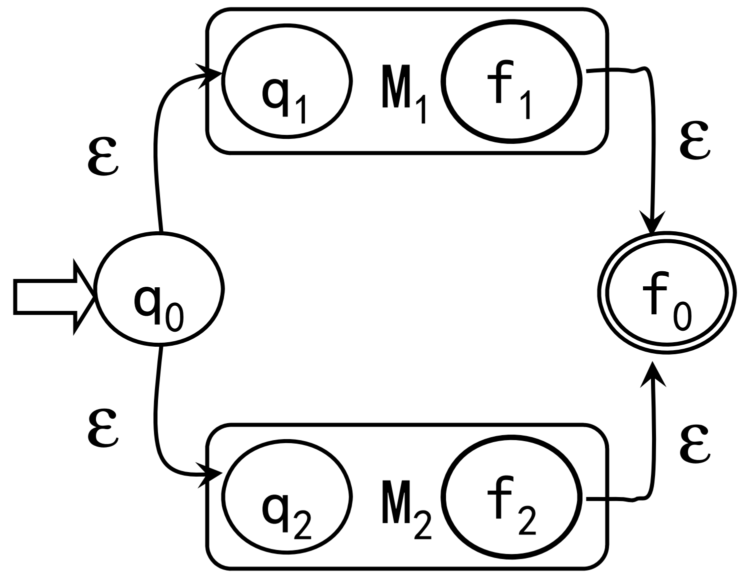
为正规式构造NFA

- $r=r_1|r_2$
- $r=r_1.r_2$
- $r=r_1{}^*$

- 令$M=<S_1 \cup S_2 \cup \{q_0,f_0\}, \Sigma_1 \cup \Sigma_2, \delta, q_0, \{f_0\}>$，其中δ定义如下(M的状态转换如右图)：
  - (a) $\delta(q_0, \varepsilon)=\{q_1,q_2\}$
  - (b) $\delta(q,a)= \delta_1(q,a)$, 当$q\in S_1-\{f_1\}$, $a\in \Sigma_1 \cup \{\varepsilon\}$
  - (c) $\delta(q,a)= \delta_2(q,a)$, 当$q\in S_2-\{f_2\}$, $a\in \Sigma_2 \cup \{\varepsilon\}$
  - (d) $\delta(f_1,\varepsilon)=\delta(f_2,\varepsilon)=\{f_0\}$。
- $L(M)=L(M_1) \cup L(M_2)$

$=L(r_1) \cup L(r_2)=L(r_1|r_2)= L(r)$

$r=r_1|r_2$



M

为正规式构造NFA

- $r=r_1|r_2$
- $r=r_1.r_2$
- $r=r_1{}^*$

- 情形1： $r=r_1|r_2$ 结论成立
- 情形2： $r=r_1.r_2$, 设M_i同情形1($i=1,2$)
  - 令$M=<S_1\cup S_2, \Sigma_1\cup\Sigma_2, \delta, q_1, \{f_2\}>$ ，其中δ定义如下(M的状态转换如右图)：

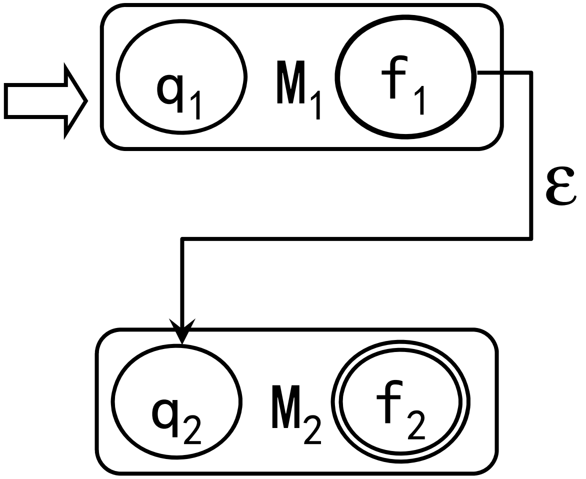
(a) $\delta(q,a)= \delta_1(q,a)$, 当$q\in S_1-\{f_1\}$, $a\in\Sigma_1\cup\{\varepsilon\}$

(b) $\delta(q,a)= \delta_2(q,a)$, 当$q\in S_2$, $a\in\Sigma_2\cup\{\varepsilon\}$

(c) $\delta(f_1,\varepsilon)=\{q_2\}$

  - $L(M)=L(M_1)L(M_2)$ $=L(r_1)L(r_2)=L(r_1.r_2)=L(r)$

$r=r_1r_2$



M

为正规式

1. 对任何FA M，都存在一个正规式r，使得L(r)=L(M)。
2. 对任何正规式r，都存在一个FA M，使得L(M)=L(r)。

- $r=r_1|r_2$
- $r=r_1.r_2$
- $r=r_1{}^*$

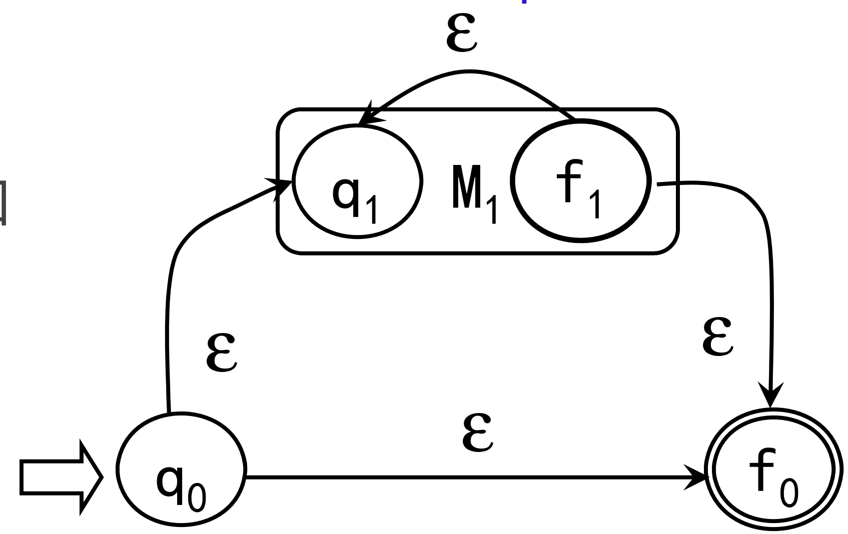
- 情形1：$r=r_1|r_2$ 结论成立
- 情形2：$r=r_1r_2$ 结论成立
- 情形3：$r=r_1{}^*$。设M_1同情形1
  - 令$M=<S_1\cup\{q_0, f_0\}, \Sigma_1, \delta, q_0, \{f_0\}>$，其中$q_0, f_0 \notin S_1$，δ定义如下(M的状态转换如右图)：

(a) $\delta(q_0,\varepsilon)=\delta(f_1,\varepsilon)=\{q_1, f_0\}$

(b) $\delta(q,a)= \delta_1(q, a)$, 当$q\in S_1-\{f_1\}$, $a\in\Sigma_1\cup\{\varepsilon\}$

  - $L(M) = L(M_1)^* = L(r_1)^* = L(r_1{}^*) = L(r)$

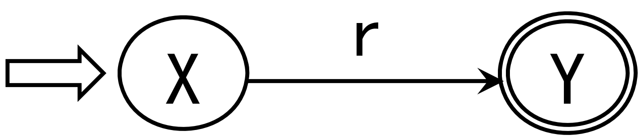
$r=r_1{}^*$



M

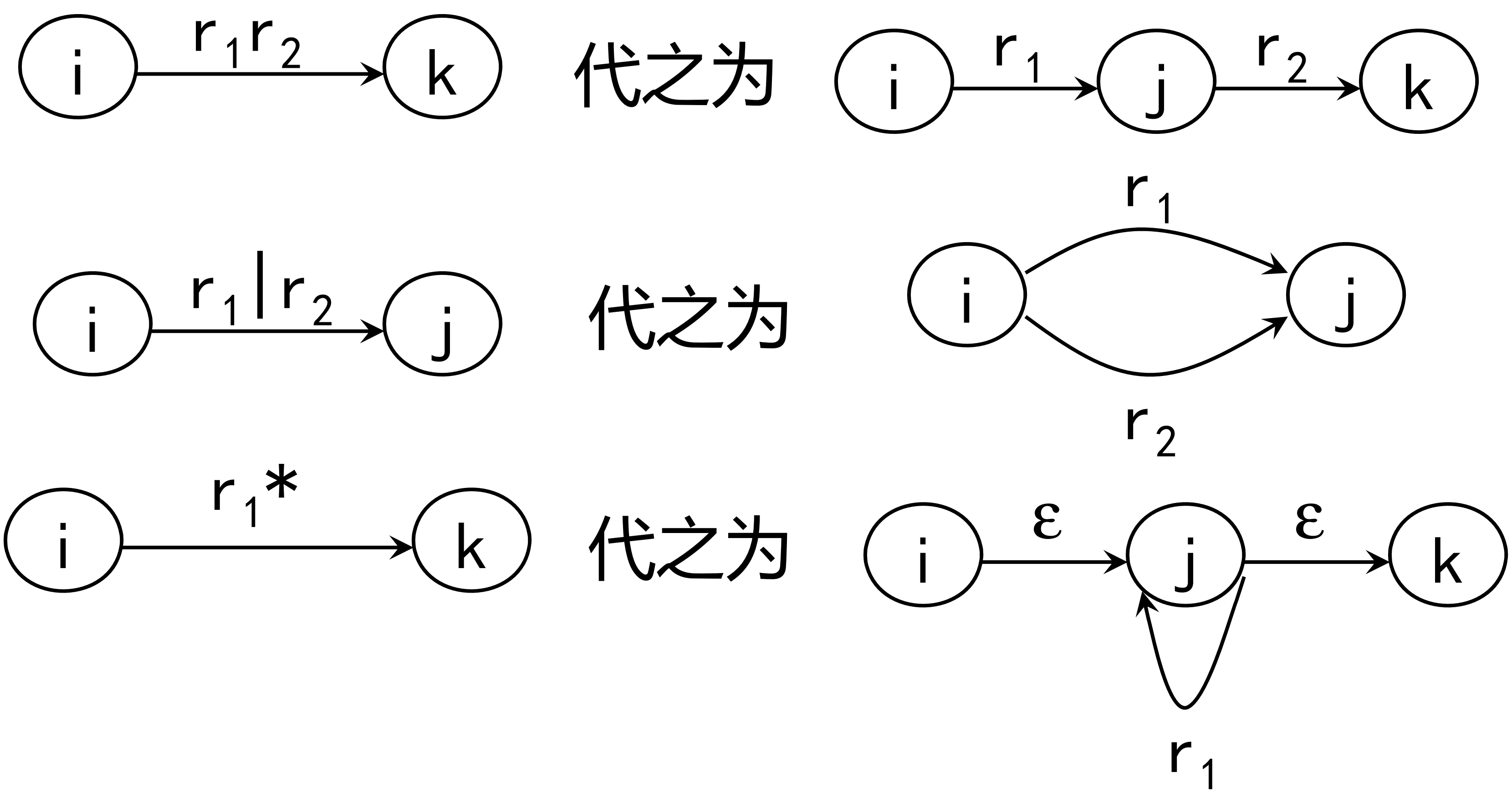
为正规式构造NFA

- 上述证明过程实质上是一个将正规表达式转换为有限自动机的算法
- 构造Σ上的NFA M' 使得 L(r)=L(M')
  - 首先，把r表示成



为正规式构造NFA

- 按下面的三条规则对r进行分裂

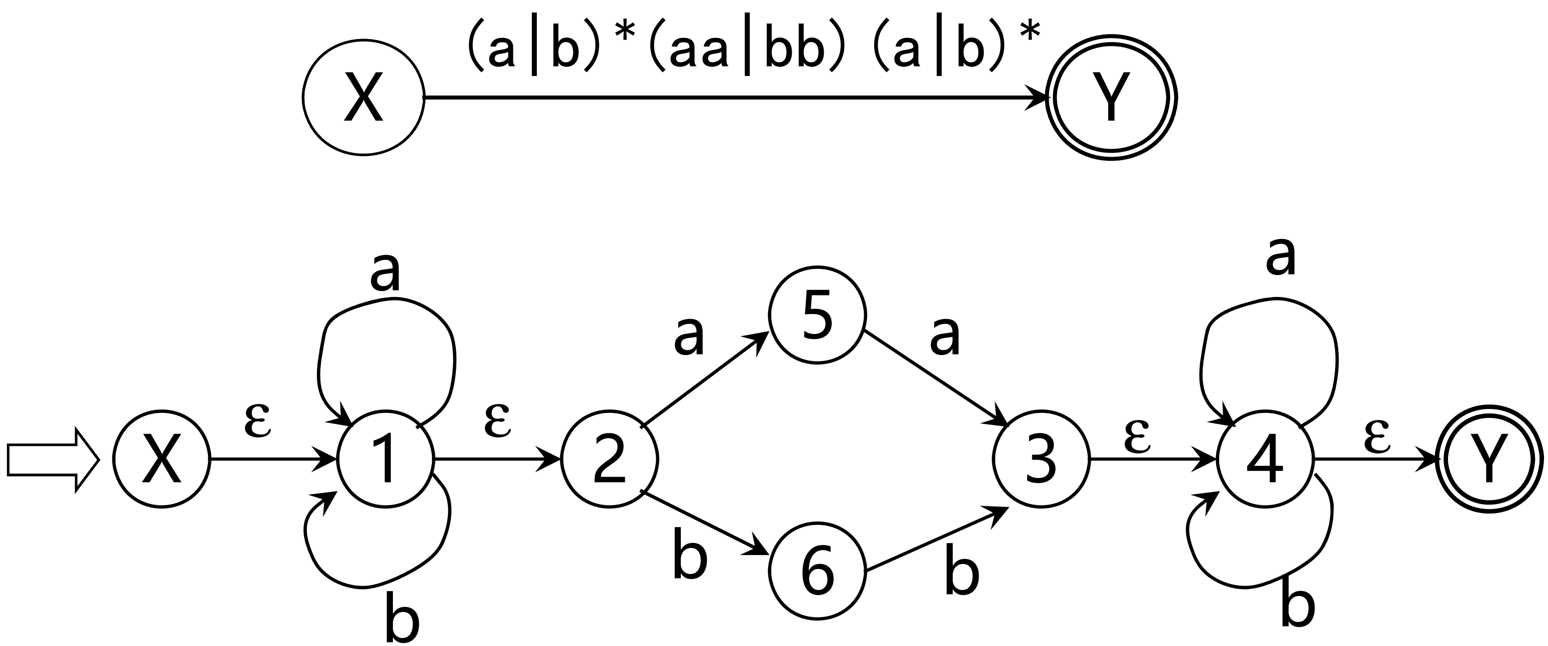


为正规式构造NFA

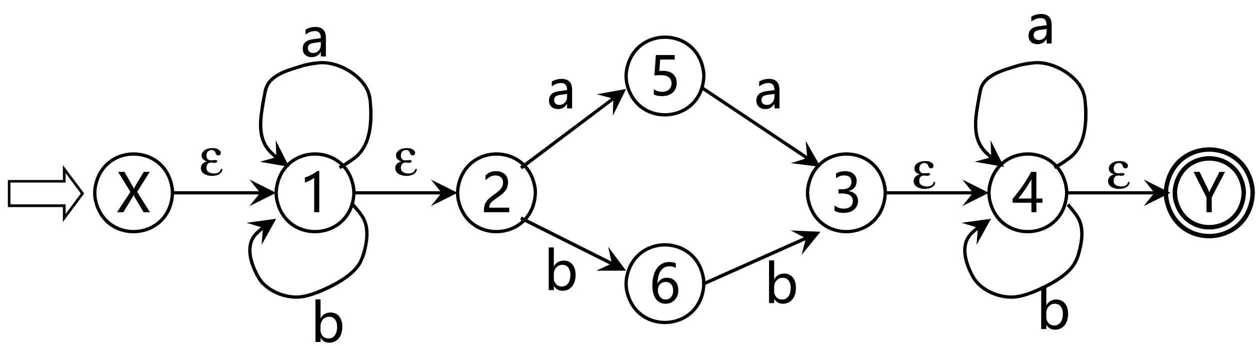
- 逐步把这个图转变为每条弧只标记为Σ上的一个字符或ε，最后得到一个NFA M’，显然L(M’)=L(r)

为正规式构造NFA

- (a|b)*(aa|bb)(a|b)*



为正规式构造NFA

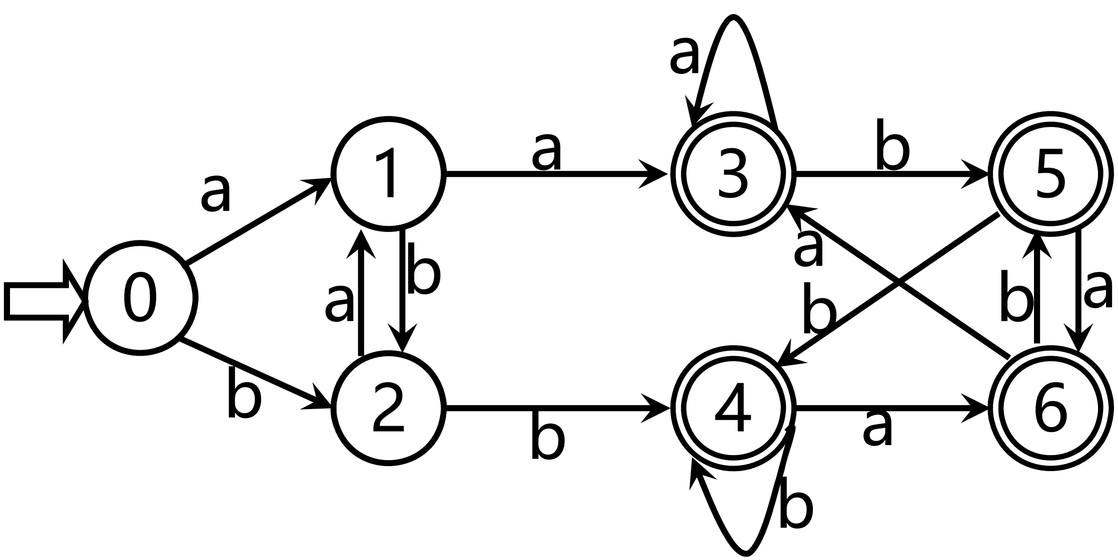


I	I_a	I_b
ε-closure({X})= {X, 1, 2}	{1, 5, 2}	{1, 6, 2}
{1, 5, 2}	{1, 3, 5, 2, 4, Y}	{1, 6, 2}
{1, 6, 2}	{1, 5, 2}	{1, 3, 6, 2, 4, Y}
{1, 3, 5, 2, 4, Y}	{1, 3, 5, 2, 4, Y}	{1, 6, 4, 2, Y}
{1, 3, 6, 2, 4, Y}	{1, 5, 4, 2, Y}	{1, 3, 6, 2, 4, Y}
{1, 6, 4, 2, Y}	{1, 5, 4, 2, Y}	{1, 3, 6, 2, 4, Y}
{1, 5, 4, 2, Y}	{1, 3, 5, 2, 4, Y}	{1, 6, 4, 2, Y}

为正规式构造NFA

I	I_a	I_b
{X, 1, 2}	{1, 5, 2}	{1, 6, 2}
{1, 5, 2}	{1, 3, 5, 2, 4, Y}	{1, 6, 2}
{1, 6, 2}	{1, 5, 2}	{1, 3, 6, 2, 4, Y}
{1, 3, 5, 2, 4, Y}	{1, 3, 5, 2, 4, Y}	{1, 6, 4, 2, Y}
{1, 3, 6, 2, 4, Y}	{1, 5, 4, 2, Y}	{1, 3, 6, 2, 4, Y}
{1, 6, 4, 2, Y}	{1, 5, 4, 2, Y}	{1, 3, 6, 2, 4, Y}
{1, 5, 4, 2, Y}	{1, 3, 5, 2, 4, Y}	{1, 6, 4, 2, Y}

I	a	b
0	1	2
1	3	2
2	1	4
3	3	5
4	6	4
5	6	4
6	3	5



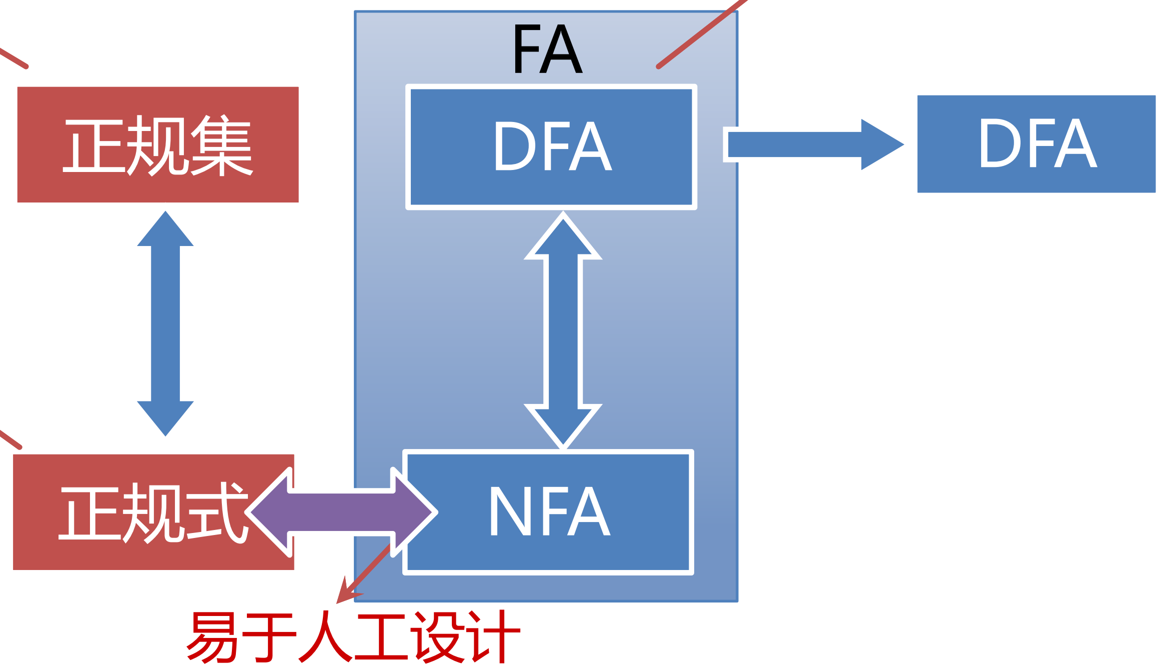
单词符号	种别编码	助忆符	内码值
DIM	1	$DIM	-
IF	2	$IF	-

正规式、正

```
curState = 初态
GetChar();
while( stateTrans[curState][ch]有定义){
    //存在后继状态，读入、拼接
    Concat();
    //转换入下一状态，读入下一字符
    curState= stateTrans[curState][ch];
    if curState是终态 then 返回strToken中的单词
    GetChar( );
}
```

DIM,IF, DO,STOP,END
number, name, age
125, 2169
…

DIM
IF
DO
STOP
END
letter(letter|digit)*
digit(digit)*



小结

- 正规式与有限自动机的等价性
  - 对任何FA M，都存在一个正规式r，使得L(r)=L(M)
    - 为NFA构造正规式
  - 对任何正规式r，都存在一个FA M，使得L(M)=L(r)
    - 为正规式构造NFA