第3章 形式语言与自动机

Part II 自动机

确定有穷自动机的化简

- 一个有穷自动机是化简了的⇔它没有多余状态并且它的状态中没有两个是互相等价的
- 一个有穷自动机可以通过消除多余状态和合并等价状态而转换成一个最小的与之等价的有穷自动机
- 所谓有穷自动机的多余状态，是指这样的状态：从自动机的开始状态出发，任何输入串也不能到达的那个状态；或者从这个状态没有通路到达终态

DFA最小化就是寻求最小状态DFA

最小状态DFA的含义:

没有多余状态（死状态）

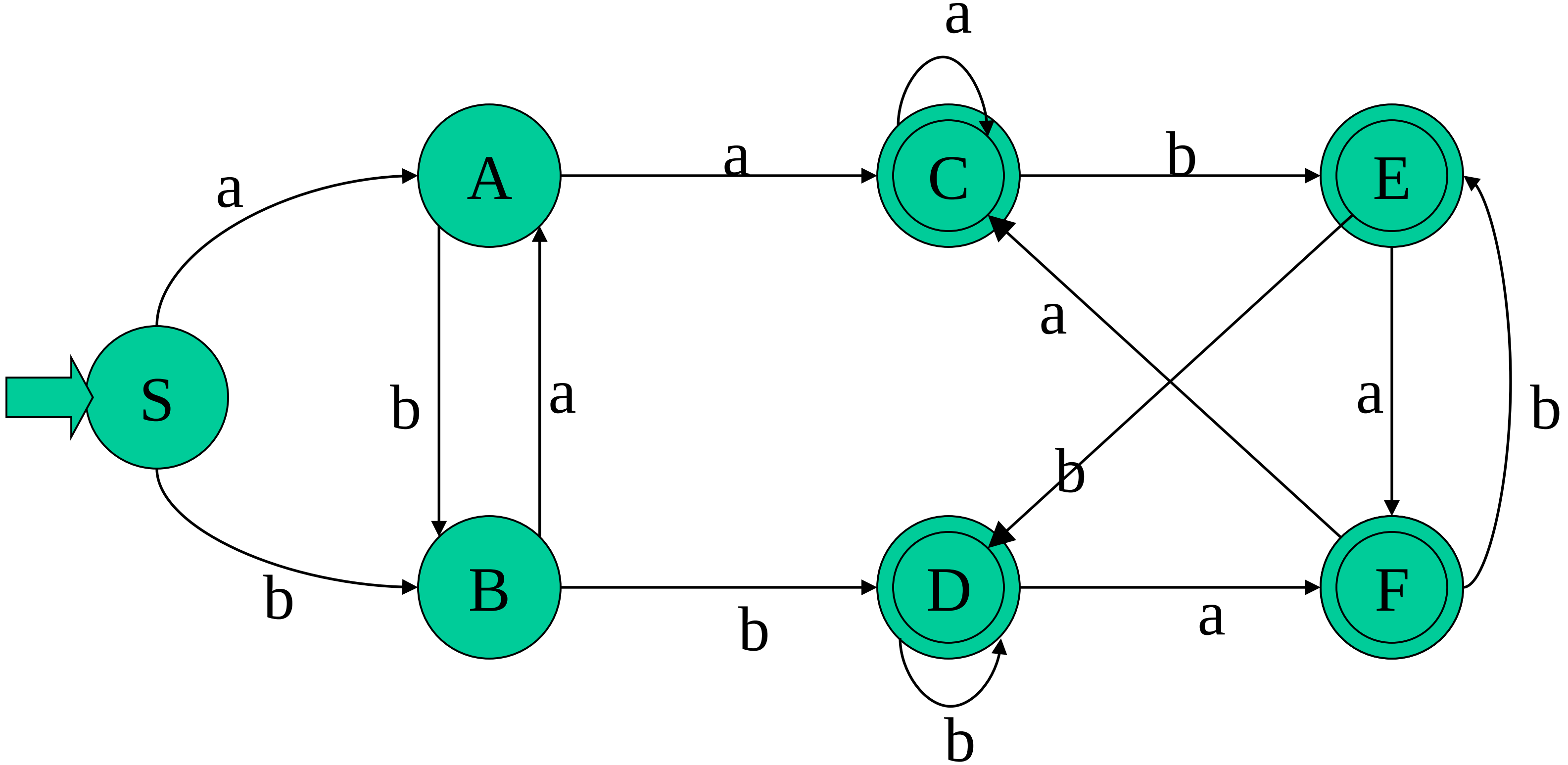
没有两个状态是互相等价（不可区别）

两个状态s和t可区别：不满足

兼容性——同是终态或同是非终态

传播性——从s出发读入某个a($a \in \Sigma$)和从t出发读入某个a到达的状态等价

C和D同是终态,读入a到达C和F, C和F同是终态,
C和F读入a都到达C,读入b都到达E. **C和D等价**



最小状态DFA

对于一个DFA M =(K,∑, f, k_0, k_t)，存在一个最小状态DFA M’ =(K’,∑, f’, k_0’, k_t’)，使L(M’)=L(M)

结论: 接受L的最小状态有穷自动机（不计同构）是唯一的

“分割法”

DFA最小化算法的核心思想：

把一个DFA的状态分成一些不相交的子集,使得任何不同的两子集的状态都是可区别的,而同一子集中的任何两个状态都是等价的

算法假定每个状态射出的弧都是完全的,否则,引入一个新状态,叫死状态,该状态是非终止状态,将不完全的输入弧都射向该状态,对所有输入,该状态射出的弧还回到自己

DFA的最小化算法

DFA M =（K,$\sum$,f,k_0,k_t),最小状态DFA M':

1.构造状态的一初始划分$\prod$**:**

终态k_t和非终态K- k_t两组(group)

2.对$\prod$施用**过程PP**构造新划分$\prod_{new}$

3.如$\prod_{new}$ =$\prod$,则令 $\prod_{final}$=$\prod$ 并继续步骤4,否则 $\prod$:=$\prod_{new}$重复2

4.为$\prod_{final}$中的每一组选一代表,这些代表构成M’的状态.若k是一代表且f(k,a)=t,令r是t组的代表,则M’中有一转换f’(k,a)=r

M’的开始状态是含有S_0那组的代表;M’ 的终态是含有F那组的代表

5.去掉M’中的死状态

过程PP：Construction of $\prod$new

For each group G of $\prod$ ***do***

begin

1.Partiton G into subgroups such that two states s and t of G are in the same subgroups if and only if for all input symbols a, states s and t have transitions on a to states in the same group of $\prod$;/*at worst, a state will be in a subgroup by itself*/

2.replace G in $\prod_{new}$ by the set of all subgroups formed

end

DFA的最小化：例子

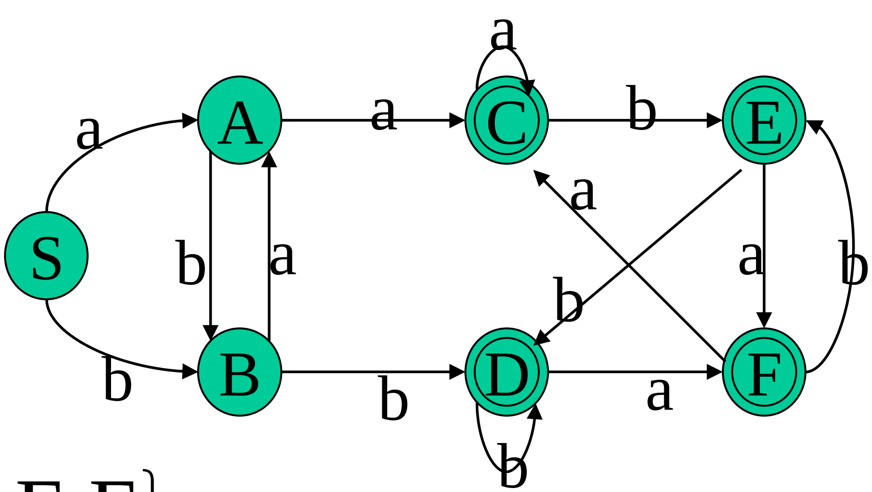
$\Pi 0$: {S,A,B}

{C,D,E,F}

$\Pi 1$: {S,A,B}

a

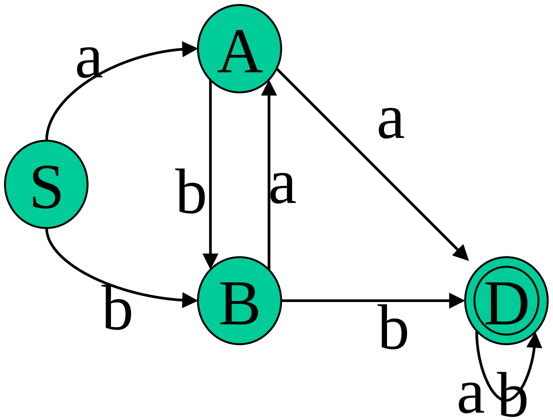
{A}　{S,B}　{C,D,E,F}



$\Pi 2$:

b

{ S } {B}



DFA的最小化算法

1. Construct an initial partition $\prod$ of the set of states with two groups: the accepting states F and the non-accepting states S-F.

2. Apply the procedure PP. to $\prod$ to construct a new partition $\prod_{new}$.

3. If $\prod_{new} = \prod$, let $\prod_{final} = \prod$ and continue with step (4). Otherwise, repeat step(2) with $\prod := \prod_{new}$.

4. Choose one state in each group of the partition $\prod_{final}$ as the representative for the group. The representatives will be the states of the reduced DFA M'. Let s be a representative state, and suppose on input a there is a transition from s to r on a.

Let the start state of M’ be the representative of the group containing the start state s_0 of M, and let the accepting states of M’ be the representatives that are in F. Note that each group of $\prod_{final}$ either consists only of states in F or has no states in F.

5. If M’ has a dead state, that is, a states d that is no accepting and that has transition to itself on all input symbols, then remove d from M’, also remove any states not reachable from the start state. Any transition to d from other states become undefined.

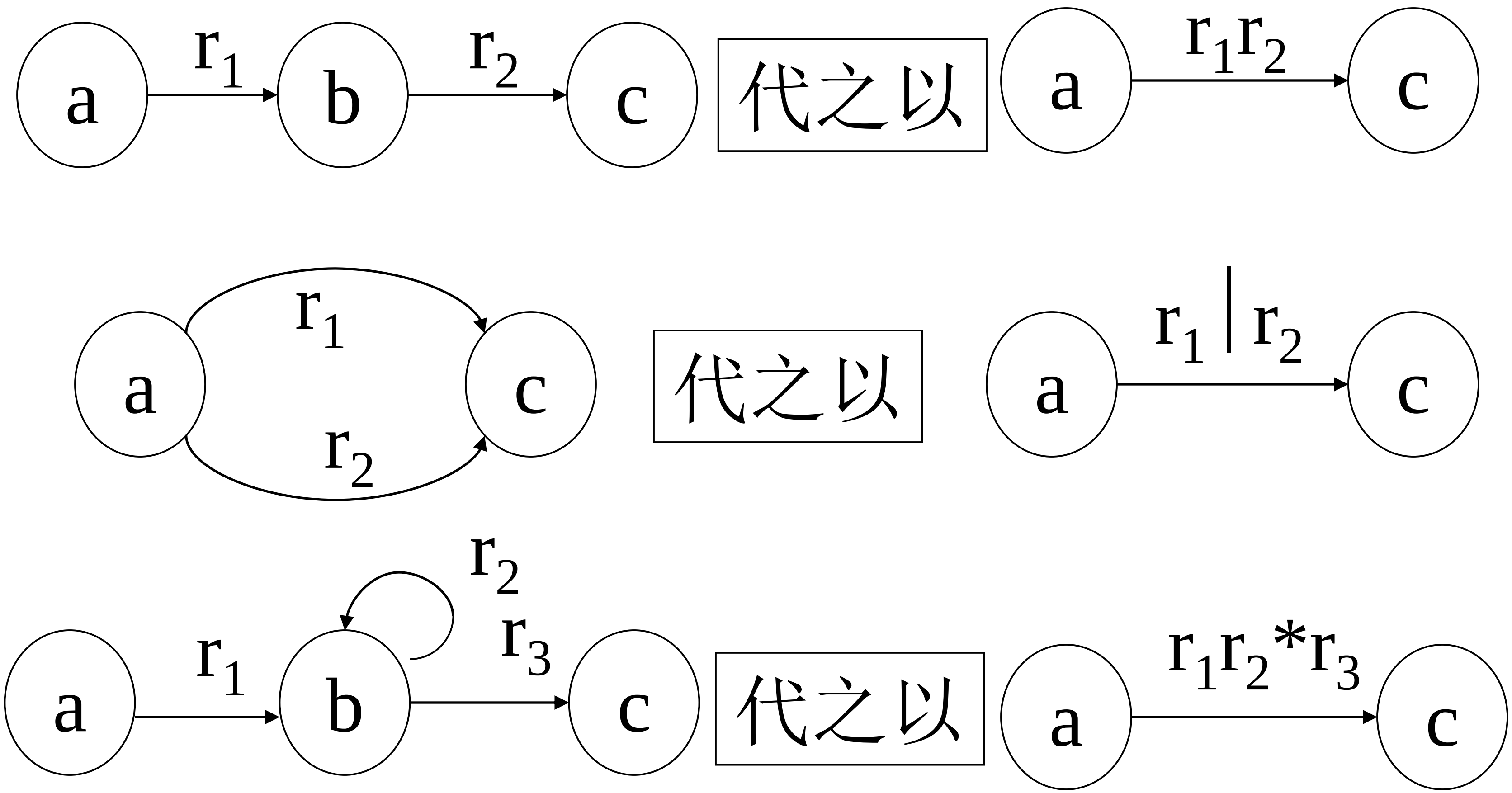
正规表达式与有限自动机的等价性

- **定理** 设 r 是 ∑ 上一个正规表达式，则存在一个 FA M 接受 L(r)，反之亦然

有限自动机=>正规表达式

- 将转换图的概念拓广，每条弧上可以用一个正规式标记。**首先**，在m的转换图上加进x, y两个结点。从x用ε弧连接到m的所有初态结点，从m的所有接受态结点用ε弧连接到y，从而构成一个新的NFA m’，L(m)=L(m’)
- **然后**，逐步消去NFA m’中的状态结点，直至剩下x,y为止。在消结的过程中，逐步用正规式标记弧。消结的过程是直观的，只需反复使用下面的替换规则：

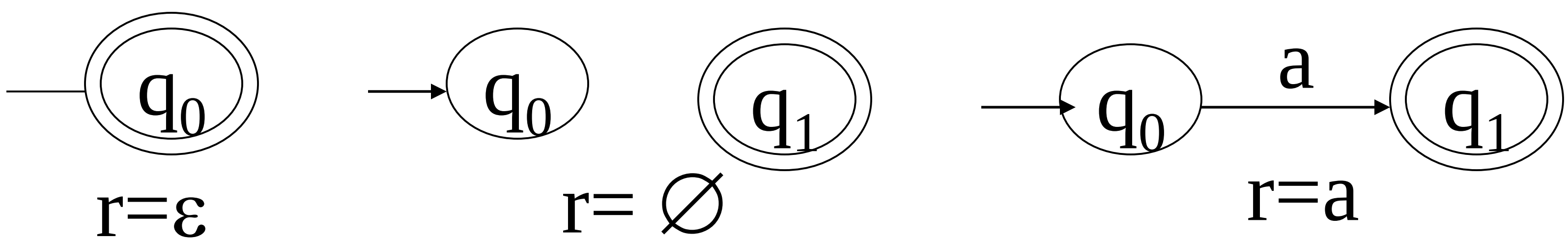
替换规则


a
r1
b
r2
c
代之以
a
r1r2
c
a
r1
c
r2
代之以
a
r1|r2
c
r2
a
r1
b
r3
c
代之以
a
r1r2*r3
c

正规表达式=>有限自动机

- 归纳法

设r具有零个运算，则或r=ε或r= ∅或r=a∈ ∑



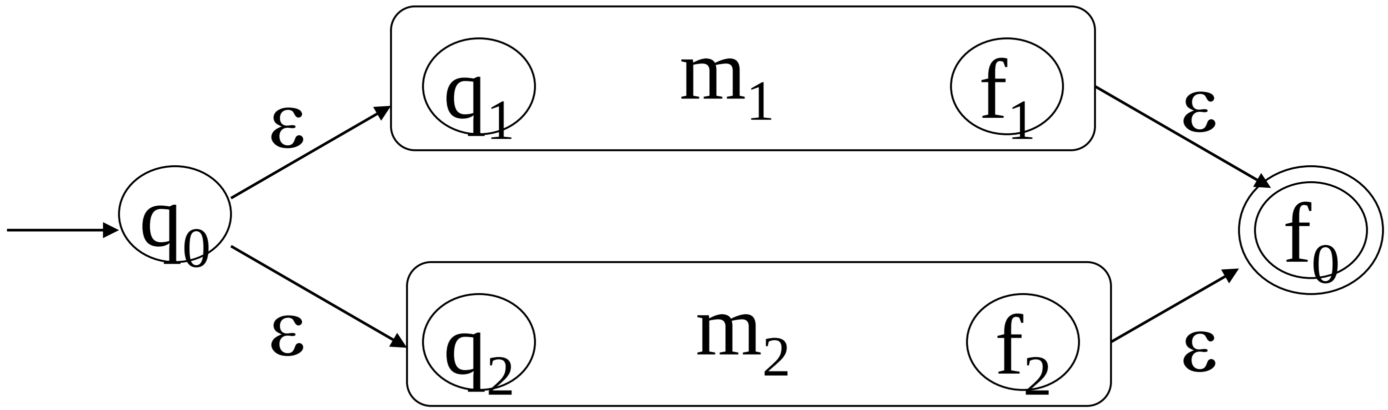
假设结论对少于i ($i \geq 1$)个运算的正规表达式r成立。当r有i个运算时，有三种情况：

情况1 $r=r_1|r_2$ **情况2** $r=r_1r_2$ **情况3** $r=r_1*$

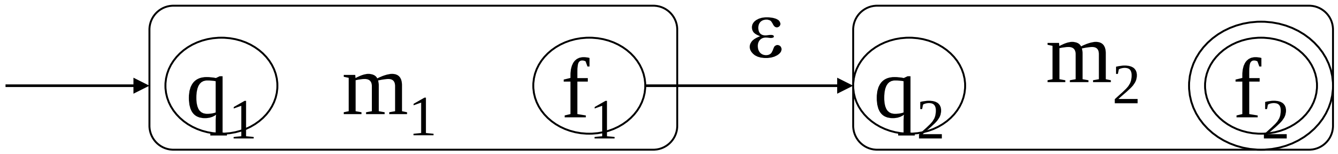
有 $m_1=(\Sigma_1,Q_1,q_1,F_1,\delta_1)$,

$m_2=(\Sigma_2,Q_2,q_2,F_2,\delta_2)$,

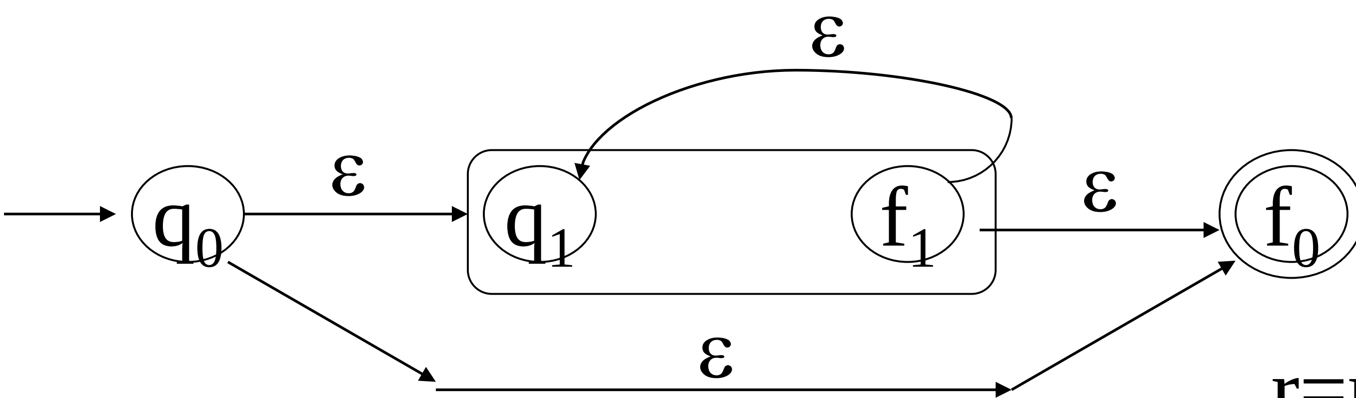
且$L(m_1)=L(r_1)$, $L(m_2)=L(r_2)$，由m_1和m_2构造m,使得 $L(m)=L(r)$.构造方法图示如下：



$r = r_1 | r_2$



$r=r_1r_2$

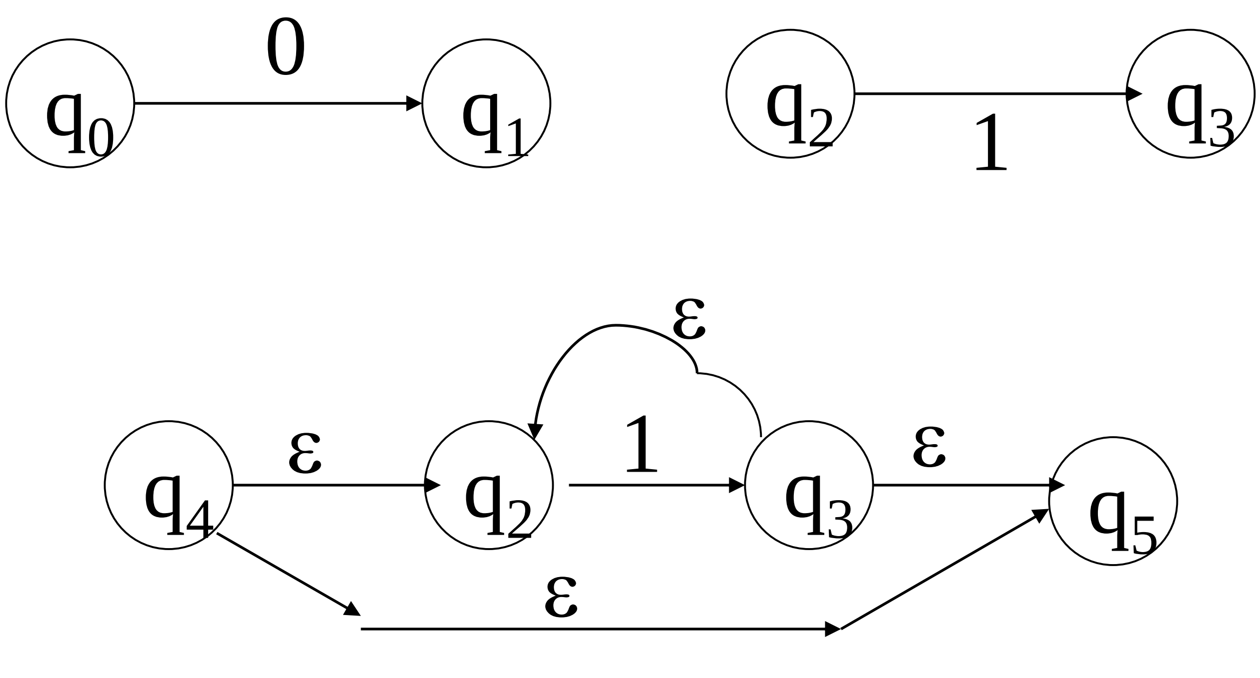


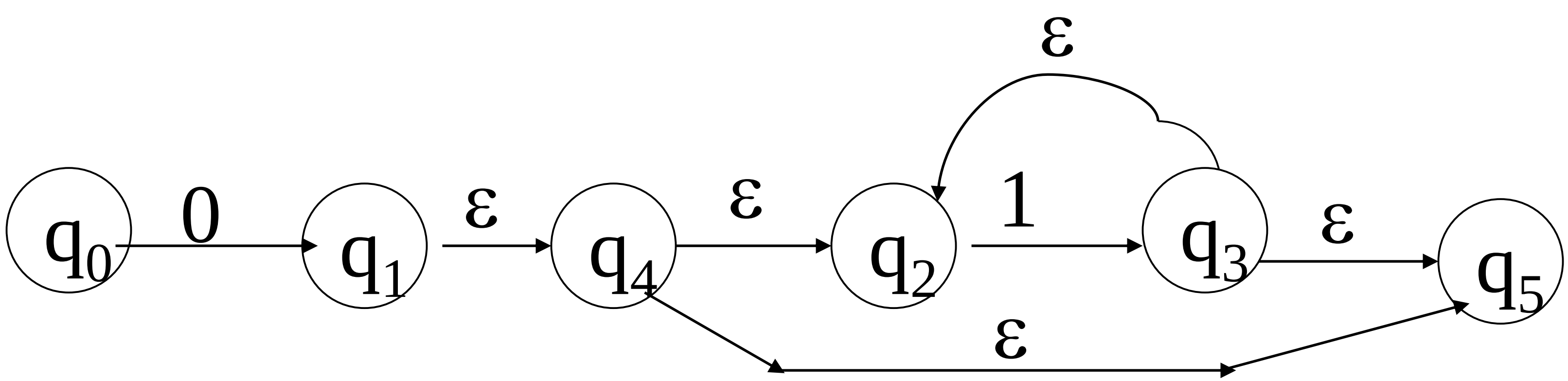
$r=r_1{}^*$

例：构造与下列正规式

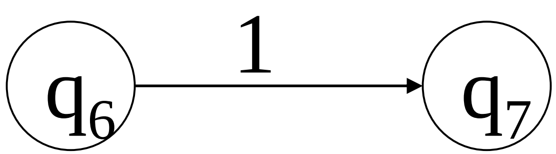
$r=01^{*}|1$

等价的有限自动机

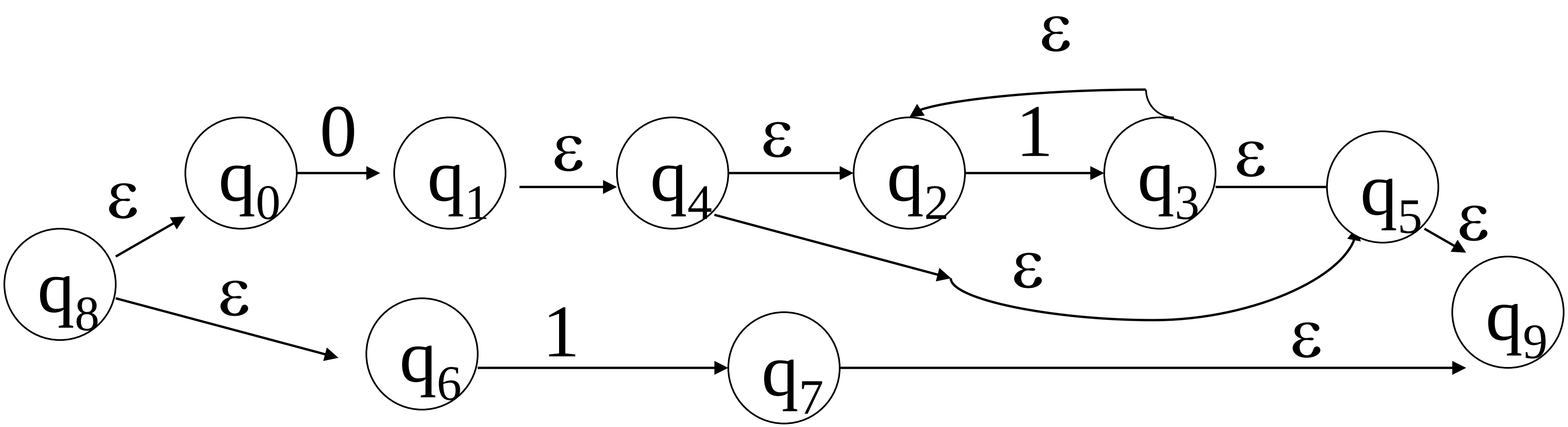




ε
q0
0
q1
ε
q4
ε
q2
1
q3
ε
q5
ε

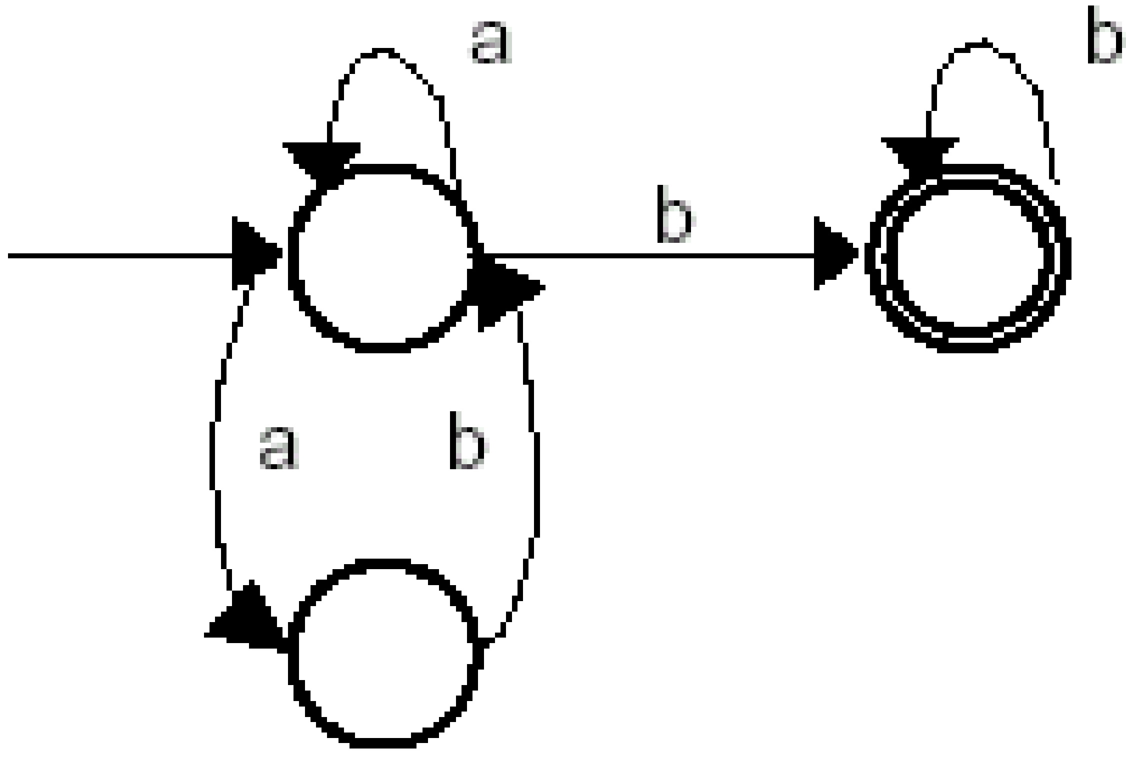


q6
1
q7



ε
q8
ε
q0
0
q1
ε
q4
ε
q2
1
q3
ε
q5
ε
ε
q6
1
q7
ε
q9

R=(a | ab)* b b*



正规文法与有限自动机的等价性

定理： 对于每一个右线性正规文法或左线性正规文法G，都存在一个FA m，使L(m)=L(G)

右线性正规文法

给定右线性正规文法$\mathbf{G=(V_T,V_N, S, P)}$, 设 $\mathbf{f \notin V_N}$, 令$\mathbf{m=(K , V_T , \delta, S, Z)}$, 其中, $\mathbf{Z= \{f\}}$ $\mathbf{K= V_N \cup \{f\}}$, 转移函数 δ 定义如下：

(a) $\mathbf{A \to a,} \qquad \mathbf{\delta(A,a)=f}$

(b) $\mathbf{A \to aA_1 \mid aA_2 \mid \ldots \mid aA_n}$

$$\mathbf{\delta(A,a)= \{A_1,A_2 ,\ldots , A_n\}}$$

左线性正规文法

给定左线性正规文法G=(V_T, V_N, S, P)，设 $q_0 \notin V_N$，令m=(K , V_T , δ, q_0 ,{S })，其中，K= $V_N \cup \{ q_0 \}$，转移函数δ定义如下：

(a) $A \rightarrow a$, $\delta(q_0,a)=A$

(b) $A_1 \rightarrow Aa$, $A_2 \rightarrow Aa$,…, $A_n \rightarrow Aa$

$\delta(A,a)= \{A_1, A_2 ,\ldots , A_n\}$

G[S]:

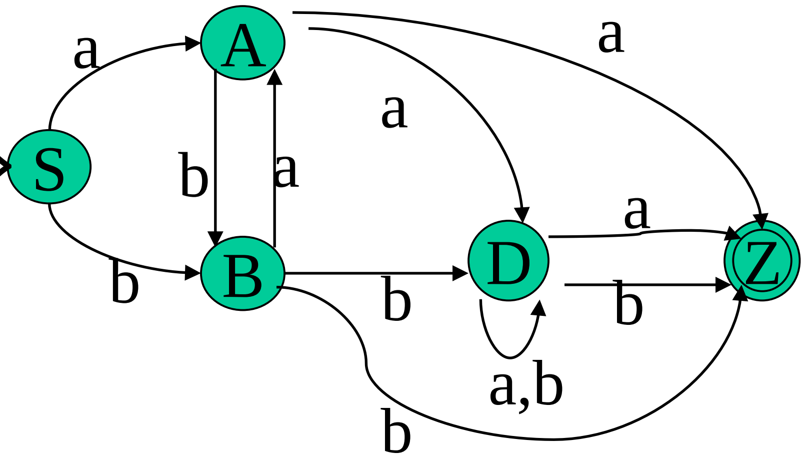
S→aA|bB

A→bB|aD|a

B→aA|bD|b

D→aD|bD|a|b

⇒



- **定理** 对于每一个FA m, 都存在一个右线性正规文法G和一个左线性正规文法G’, 使L(G)=L(G’)=L(m)

右线性正规文法的构造方法

给定m=($\sum$, V_N , S, Z, δ)，则G=($\sum$,V_N, S, P),

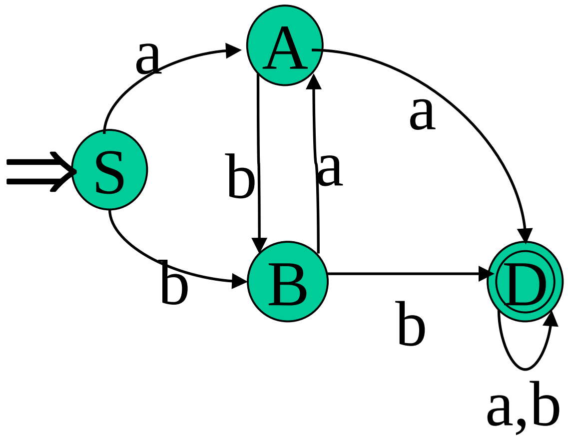
其中P的定义如下：对任何a $\in \sum$及A,B $\in V_N$,

若有δ(A,a)=B, 则

(a) B $\notin$Z, $A \to aB$

(b) B $\in$Z $A \to a \mid aB$

若S$\in$Z, 则$S \to \varepsilon$



G[S]:

$S \rightarrow aA|bB$

$A \rightarrow bB|aD|a$

$B \rightarrow aA|bD|b$

$D \rightarrow aD|bD|a|b$

构造左线性正规文法

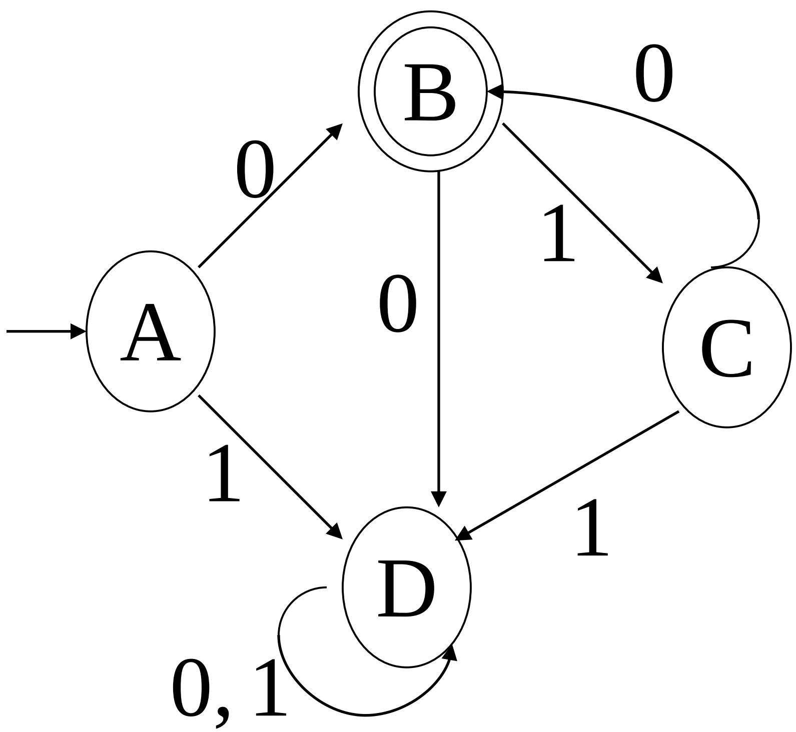
- 构造左线性正规文法, P的定义如下：
- 对任何$a \in \sum$及$A_1, A_2 \in V_N$, 有$\delta(A_1, a) = A_2$, 则：

(a) A_1是初态, $A_2 \rightarrow a$

(b) A_1不是初态, $A_2 \rightarrow A_1 a$

例: DFA m $\Rightarrow$右线性正规文法G

$\Rightarrow$ NFA m'$\Rightarrow$左线性正规文法G'



$\Rightarrow$

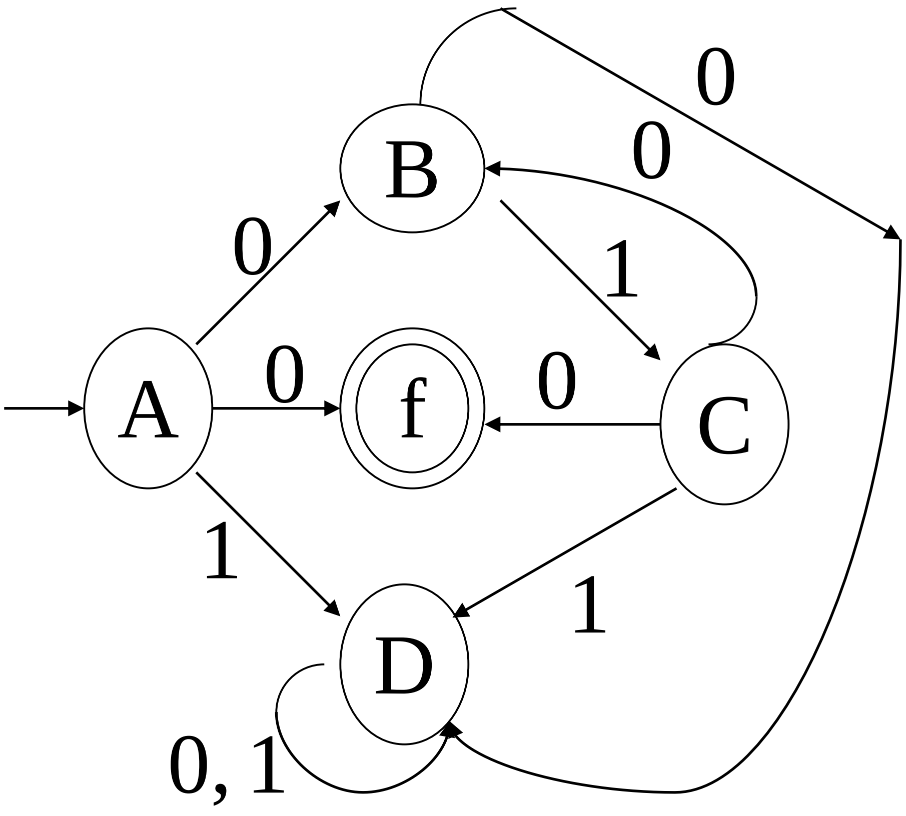
$A \to 0 \mid 0B \mid 1D$

$B \to 0D \mid 1C$

$C \to 0 \mid 0B \mid 1D$

$D \to 0D \mid 1D$

NFA m’　　　　　　左线性正规文法



$\Rightarrow$

$S \rightarrow 0 \mid C0$

$B \rightarrow 0 \mid C0$

$C \rightarrow B1$

$D \rightarrow 1 \mid B0 \mid C1$

$\quad\quad \mid D0 \mid D1$

ε-FA

1）在ε-FA中寻找ε边：A $\xrightarrow{\varepsilon}$ B,并且B没有ε边指向A，此时转2），否则转4）；

2）设状态B的直接后继状态为S1,S2,…,Sk，且

$$A \xrightarrow{\varepsilon} B \xrightarrow{a_i} S_i$$

则消除原ε边，引进新边A $\xrightarrow{a_i}$ Si，消去ε边后，如果B∈Z则A也应为终态，若A为初态，则B也应为初态。

3）重复2），直到1）中指出ε边的被消除；

4）对于有ε回路的情况，则将ε回路中的状态合并为一个结点。

Summary

- 正规文法
- 正规表达式
- FA

作业

- 4(a)
- 9

本章作业

- 1(1)(3)
- 4(a)
- 8
- 9