# 第 4 章　词法分析

**第 1 题：**

构造下列正规式相应的 DFA.

（1） 1(0|1)　*101

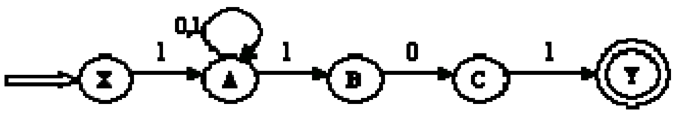
（2） 1 (1010*|1(010)*1）*0

（3） a((a|b)*|ab*a)*b

（4） b((ab)*|bb)*ab

答案：

(1) 先构造 NFA：



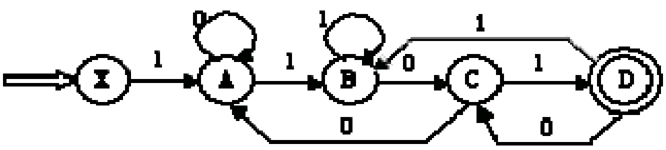

用子集法将 NFA 确定化

| . | 0 | 1 |
|---|---|---|
| X | . | A |
| A | A | AB |
| AB | AC | AB |
| AC | A | ABY |
| ABY | AC | AB |

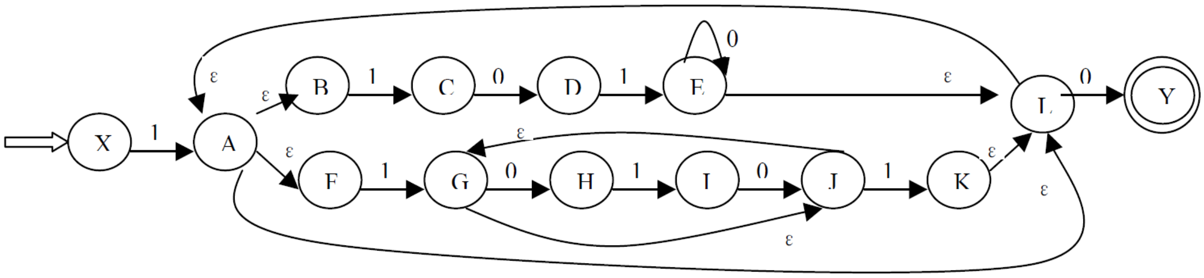
除 X，A 外，重新命名其他状态，令 AB 为 B、AC 为 C、ABY 为 D，因为 D 含有 Y（NFA 的终态），所以 D 为终态。

| . | 0 | 1 |
|---|---|---|
| X | . | A |
| A | A | B |
| B | C | B |
| C | A | D |
| D | C | B |

DFA 的状态图：：




(2)先构造 NFA：

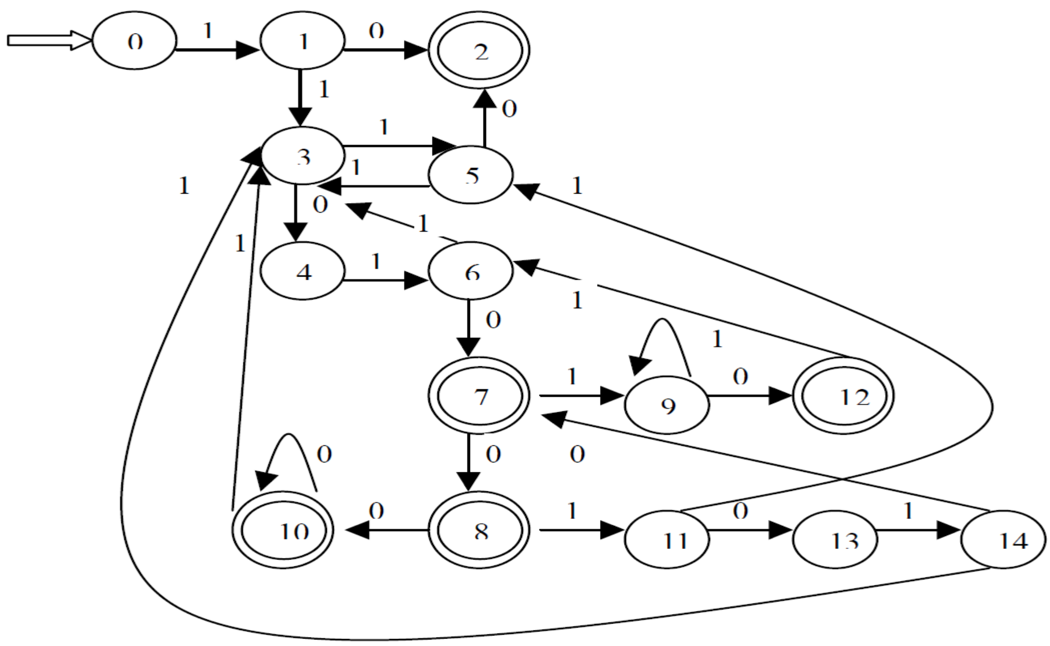


用子集法将 NFA 确定化

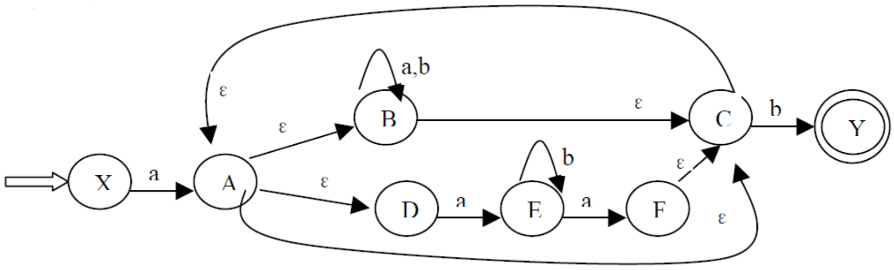
| | ε | 0 | 1 |
|---|---|---|---|
| X | X | | |
| $T_0$=X | | | A |
| A | ABFL | | |
| $T_1$= ABFL | | Y | CG |
| Y | Y | | |
| CG | CGJ | | |
| $T_2$= Y | | | |
| $T_3$= CGJ | | DH | K |
| DH | DH | | |
| K | ABFKL | | |
| $T_4$= DH | | | EI |
| EI | ABEFIL | | |
| $T_5$= ABFKL | | Y | CG |
| $T_6$= ABEFIL | | EJY | CG |
| EJY | ABEFGJLY | | |
| $T_7$= ABEFGJLY | | EHY | CGK |
| EHY | ABEFHLY | | |
| CGK | ABCFGJKL | | |
| $T_8$= ABEFHLY | | EY | CGI |
| EY | ABEFLY | | |
| CGI | CGJI | | |
| $T_9$= ABCFGJKL | | DHY | CGK |
| DHY | DHY | | |
| $T_{10}$= ABEFLY | | EY | CG |
| $T_{11}$= CGJI | | DHJ | K |
| DHJ | DHJ | | |
| $T_{12}$= DHY | | | EI |
| $T_{13}$= DHJ | | | EIK |
| EIK | ABEFIKL | | |
| $T_{14}$= ABEFIKL | | EJY | CG |

将$T_0$、$T_1$、$T_2$、$T_3$、$T_4$、$T_5$、$T_6$、$T_7$、$T_8$、$T_9$、$T_{10}$、$T_{11}$、$T_{12}$、$T_{13}$、$T_{14}$重新命名，分别用 0、1、2、3、4、5、6、7、8、9、10、11、12、13、14 表示。因为 2、7、8、10、12 中含有Y，所以它们都为终态。

| | 0 | 1 |
|---|---|---|
| 0 | | 1 |
| 1 | 2 | 3 |
| 2 | | |
| 3 | 4 | 5 |
| 4 | | 6 |
| 5 | 2 | 3 |
| 6 | 7 | 3 |
| 7 | 8 | 9 |
| 8 | 10 | 11 |
| 9 | 12 | 9 |
| 10 | 10 | 3 |
| 11 | 13 | 5 |
| 12 | | 6 |
| 13 | | 14 |
| 14 | 7 | 3 |



(3) 先构造 NFA：

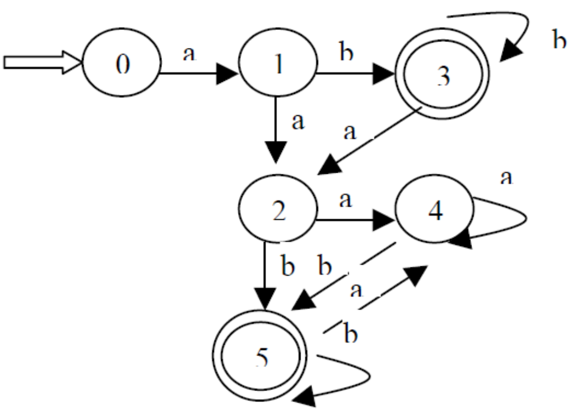



用子集法将 NFA 确定化

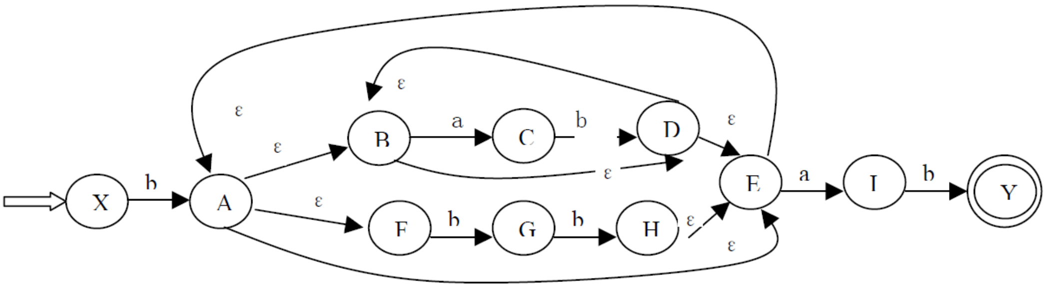
| | ε | a | b |
|---|---|---|---|
| X | X | | |
| $T_0$=X | | A | |
| A | ABCD | | |
| $T_1$=ABCD | | BE | BY |
| BE | ABCDE | | |
| BY | ABCDY | | |
| $T_2$=ABCDE | | BEF | BEY |
| BEF | ABCDEF | | |
| BEY | ABCDEY | | |
| $T_3$=ABCDY | | BE | BY |
| $T_4$=ABCDEF | | BEF | BEY |
| $T_5$=ABCDEY | | BEF | BEY |

将$T_0$、$T_1$、$T_2$、$T_3$、$T_4$、$T_5$重新命名，分别用 0、1、2、3、4、5 表示。因为 3、5 中含有Y，所以它们都为终态。

| | a | b |
|---|---|---|
| 0 | 1 | |
| 1 | 2 | 3 |
| 2 | 4 | 5 |
| 3 | 2 | 3 |
| 4 | 4 | 5 |
| 5 | | 5 |



(4) 先构造 NFA：



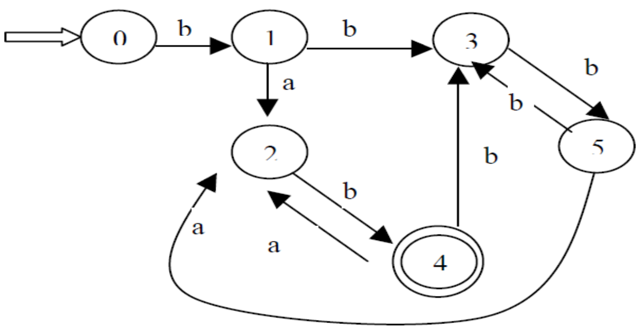

用子集法将 NFA 确定化：

| | ε | a | b |
|---|---|---|---|
| X | X | | |
| $T_0$=X | | | A |
| A | ABDEF | | |
| $T_1$=ABDEF | | CI | G |
| CI | CI | | |
| G | G | | |
| $T_2$=CI | | | DY |
| DY | ABDEFY | | |
| $T_3$=G | | | H |
| H | ABEFH | | |
| $T_4$=ABDEFY | | CI | G |
| $T_5$=ABEFH | | CI | G |

将$T_0$、$T_1$、$T_2$、$T_3$、$T_4$、$T_5$重新命名，分别用 0、1、2、3、4、5 表示。因为 4 中含有Y，所以它为终态。

| | a | b |
|---|---|---|
| 0 | | 1 |
| 1 | 2 | 3 |
| 2 | | 4 |
| 3 | | 5 |
| 4 | 2 | 3 |
| 5 | 2 | 3 |

DFA 的状态图：



**第2题** :请描述下面正规式定义的串. 字母表 {0，1}.

a) 0*(10+)*0*

b) (0|1)*(00|11) (0|1)*

c) 1(0|1)*0

答案:

a) 每个 1 至少有一个 0 跟在后边的串

b) 所有含两个相继的0或两个相继的1的串

c) 必须以 1 开头和0结尾的串

**第 3 题：** 已知 NFA＝({x,y,z}，{0,1}，M，{x}，{z})，其中：M(x,0)={z}，M(y,0)={x,y}，,M(z,0)={x,z}，M(x,1)={x}，M(y,1)=φ，M(z,1)={y}，构造相应的 DFA。

答案： 先构造其矩阵

| | 0 | 1 |
|---|---|---|
| x | z | x |
| y | x,y | |
| z | x,z | y |

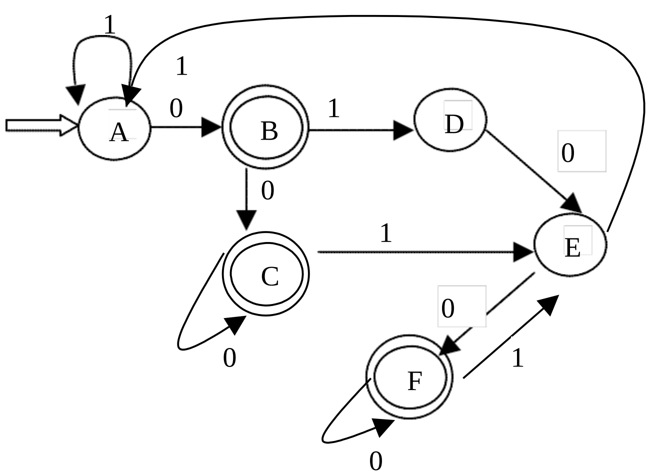
用子集法将 NFA 确定化：

| | 0 | 1 |
|---|---|---|
| x | z | x |
| z | xz | y |
| xz | xz | xy |
| y | xy | |
| xy | xyz | x |
| xyz | xyz | xy |

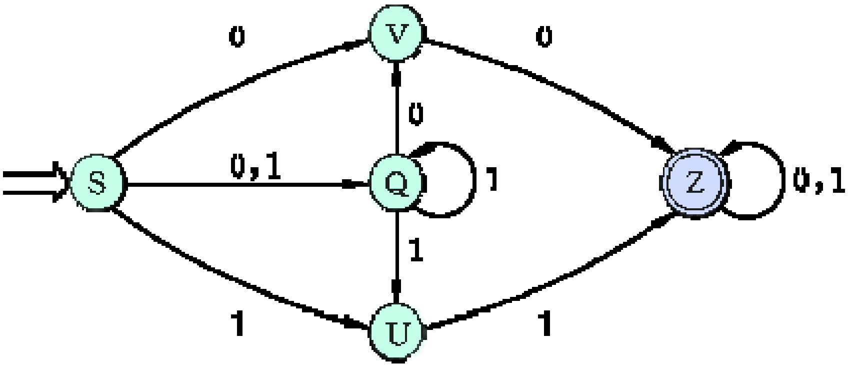
将 x、z、xz、y、xy、xyz 重新命名，分别用 A、B、C、D、E、F 表示。因为 B、C、F 中含有 z，所以它为终态。

| | 0 | 1 |
|---|---|---|
| A | B | A |
| B | C | D |
| C | C | E |
| D | E | |
| E | F | A |
| F | F | E |

DFA 的状态图：



**第 4 题**:将下图确定化:




答案:

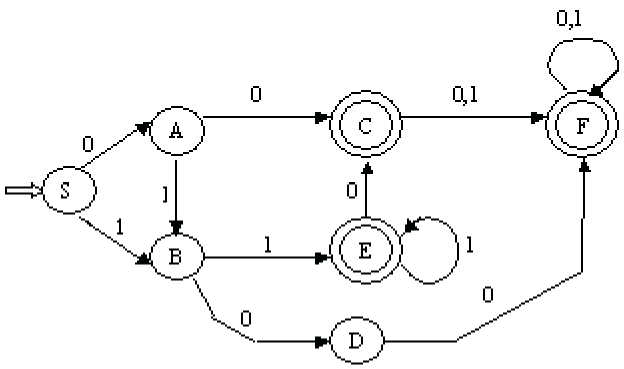
用子集法将 NFA 确定化:

| . | 0 | 1 |
|---|---|---|
| S | V | Q |
| V | V | Q |
| Q | V | QU |
| V | Z | Z |
| V | Z | . |
| QU | V | QU |
| Z | Z | Z |

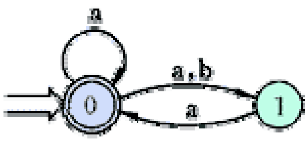
重新命名状态子集,令 VQ 为 A、QU 为 B、VZ 为 C、V 为 D、QUZ 为 E、Z 为 F。

| . | 0 | 1 |
|---|---|---|
| S | A | B |
| A | C | B |
| B | D | E |
| C | F | F |
| D | F | . |
| E | C | E |
| F | F | F |

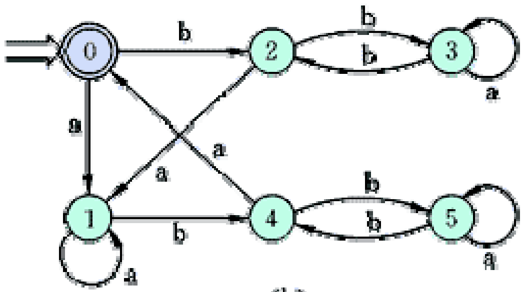
DFA 的状态图:




**第5题：** 将下图的（a）和（b）分别确定化和最小化：




(a)




(b)

答案：

初始分划得

Π0：终态组{0}，非终态组{1,2,3,4,5}

对非终态组进行审查：

{1,2,3,4,5}a ⊂{0,1,3,5}

而{0,1,3,5}既不属于{0}，也不属于{1,2,3,4,5}

∵{4} a ⊂{0}，所以得到新分划

Π1：{0}，{4}，{1,2,3,5}

对{1,2,3,5}进行审查：

∵{1,5} b ⊂{4}

{2,3} b ⊂{1,2,3,5}，故得到新分划

Π2：{0}，{4}，{1, 5}，{2,3}

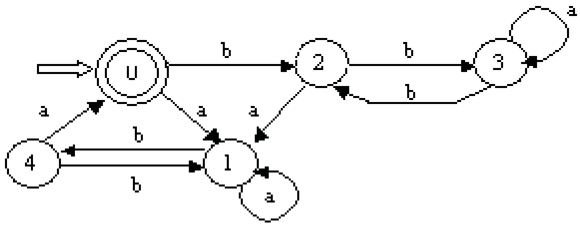
{1, 5} a ⊂{1, 5}

{2,3} a ⊂{1,3}，故状态 2 和状态 3 不等价，得到新分划

Π3：{0}，{2}，{3}，{4}，{1, 5}

这是最后分划了

最小 DFA：

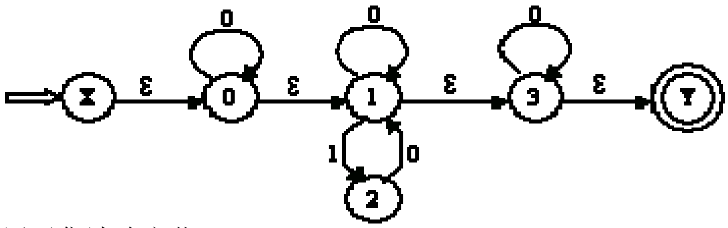



**第 6 题：**

构造一个 DFA，它接收 Σ={0,1}上所有满足如下条件的字符串：每个 1 都有 0 直接跟在右边。并给出该语言的正规式。

答案：

按题意相应的正规表达式是(0*10)*0*，或 0*(0 | 10)*0* 构造相应的 DFA，首先构造 NFA 为



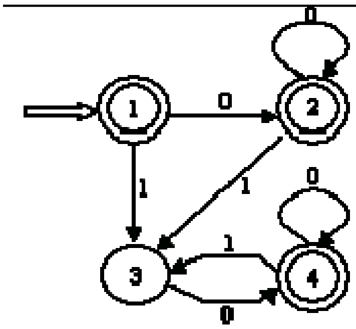

用子集法确定化：

| I | I | I |
|---|---|---|
| {X,0,1,3,Y} | {0,1,3,Y} | {2} |
| {0,1,3,Y} | {0,1,3,Y} | {2 |
| { | {1,3 | |

重新命名状态集：

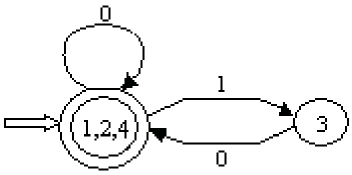
| S | 0 | 1 |
|---|---|---|
| 1 | 2 | 3 |
| 2 | 2 | 3 |
| 3 | 4 | |
| 4 | 4 | 3 |

DFA 的状态图：




可将该 DFA 最小化：

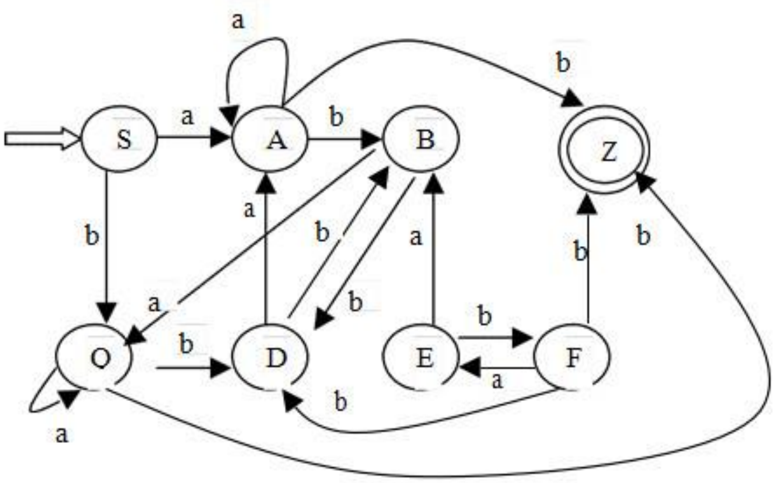
终态组为{1,2,4}，非终态组为{3}，{1,2,4}0 {1,2,4}，{1,2,4}1 {3}，所以 1,2,4 为等价状 态，可合并。




**第 7 题:给**文法 G[S]： S→aA|bQ A→aA|bB|b B→bD|aQ Q→aQ|bD|b D→bB|aA E→aB|bF F→bD|aE|b

构造相应的最小的 DFA。

答案：先构造其 NFA：




用子集法将 NFA 确定化：

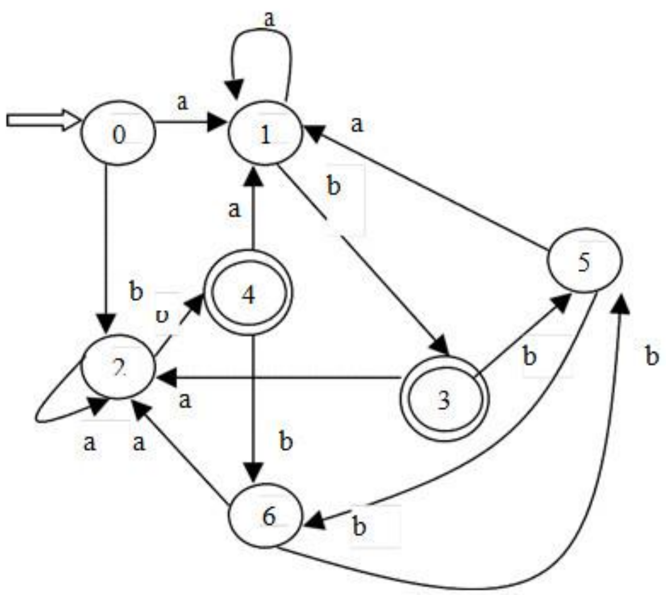
| | a | b |
|---|---|---|
| S | A | Q |
| A | A | BZ |
| Q | Q | DZ |
| BZ | Q | D |
| DZ | A | B |
| D | A | B |
| B | Q | D |

将 S、A、Q、BZ、DZ、D、B 重新命名，分别用 0、1、2、3、4、5、6 表示。因为 3、4 中含有 z，所以它们为终态。

| | a | b |
|---|---|---|

| | | |
|---|---|---|
| 0 | 1 | 2 |
| 1 | 1 | 3 |
| 2 | 2 | 4 |
| 3 | 2 | 5 |
| 4 | 1 | 6 |
| 5 | 1 | 6 |
| 6 | 2 | 5 |

DFA 的状态图：

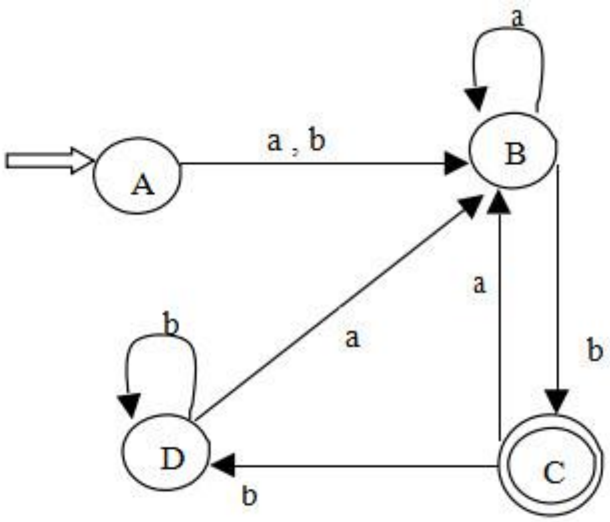



令$P_0$＝（{0,1,2,5,6}，{3,4}）用b进行分割：

$P_1$＝（{0,5, 6}，{1,2}，{3,4}）再用b进行分割：

$P_2$＝（{0}，{5, 6}，{1,2}，{3,4}）再用a、b 进行分割，仍不变。

再令｛0｝为 A，{1，2}为 B，{3，4}为 C，{5，6}为 D。 最小化为：




**第8题:**给出下述文法所对应的正规式：

S→aA|bB

A→bS|b

B→aS|a

答案：

解方程组 S 的解：

S=aA|bB

A=bS|b

B=aS|a

将 A、B 产生式的右部代入 S 中

S=abS|ab|baS|ba=（ab|ba）S|（ab|ba）

所以：S= $(ab|ba)^*(ab|ba)$

**第9题：** 考虑正规表达式 r = a*b(a | b) ，构造可以生成语言 L(r) 的一个正规文法。

答案：

$S \rightarrow a^*b(a \mid b)$

变换为 $S \rightarrow aA,\ S \rightarrow b(a \mid b),\ A \rightarrow aA,\ A \rightarrow b(a \mid b)$

变换为 $S \rightarrow aA,\ S \rightarrow bB,\ B \rightarrow (a \mid b),\ A \rightarrow aA,\ A \rightarrow bC,\ C \rightarrow (a \mid b)$

变换为 $S \rightarrow aA,\ S \rightarrow bB,\ B \rightarrow a,\ B \rightarrow b,\ A \rightarrow aA,\ A \rightarrow bC,\ C \rightarrow a,\ C \rightarrow b$

所以，一个可能的正规文法为 G[S]：

$S \rightarrow aA,\ S \rightarrow bB,\ B \rightarrow a,\ B \rightarrow b,\ A \rightarrow aA,\ A \rightarrow bC,\ C \rightarrow a,\ C \rightarrow b$

或表示为：

$S \rightarrow aA \mid bB,\ B \rightarrow a \mid b,\ A \rightarrow aA \mid bC,\ C \rightarrow a \mid b$

（适当等价变换也可以，但要作说明，即要有步骤）