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

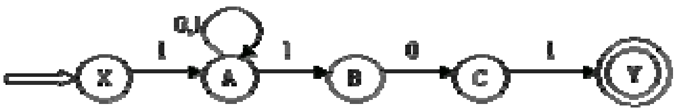
（2） 1 (1010*|1(010)*1）*0

（3） a((a|b)*|ab*a)*b

（4） b((ab)*|bb)*ab

答案：

(1) 先构造 NFA：



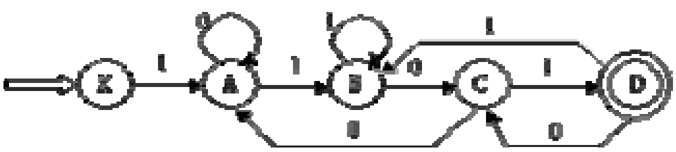
用子集法将 NFA 确定化

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

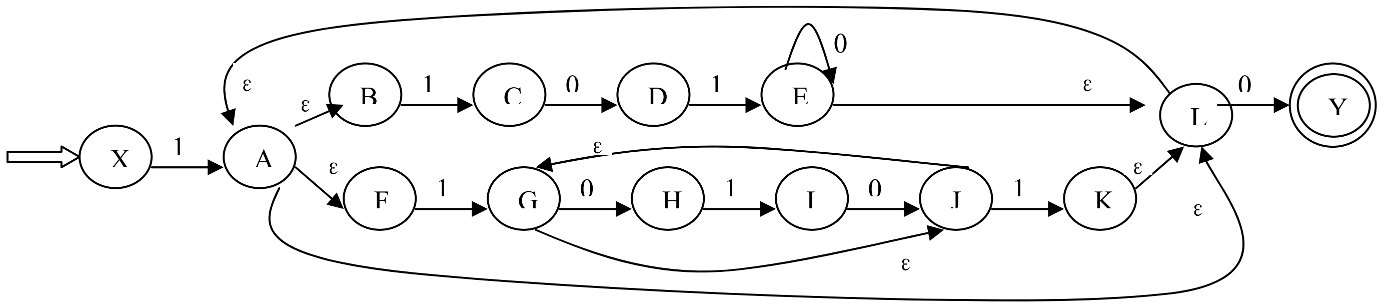
除 X，A 外，重新命名其他状态，令 AB 为 B、AC 为 C、ABY 为 D，因为 D 含有 Y（NFA 的终态），所以 D 为终态。

.	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：

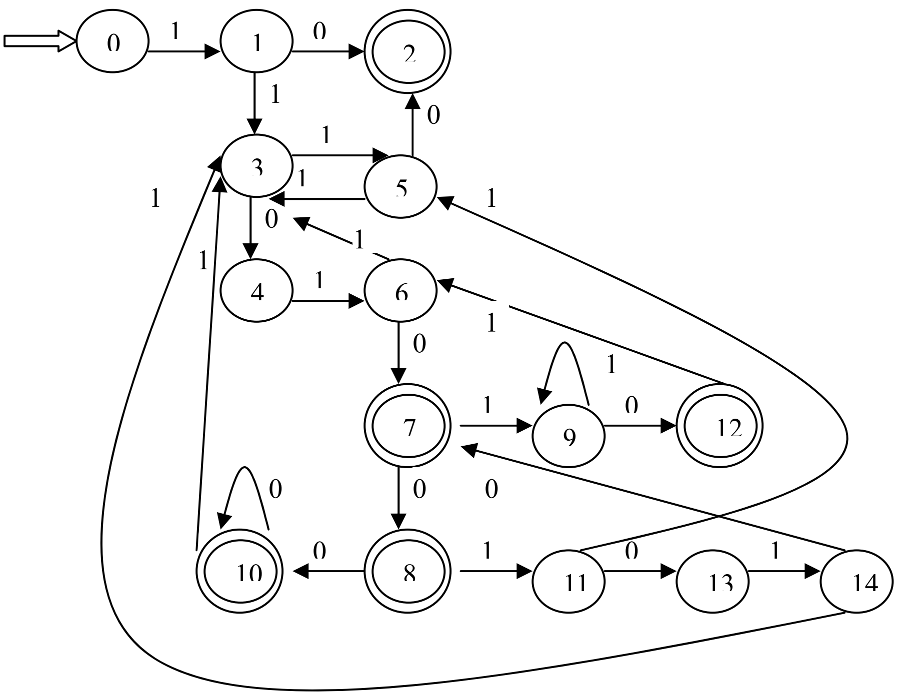


用子集法将 NFA 确定化

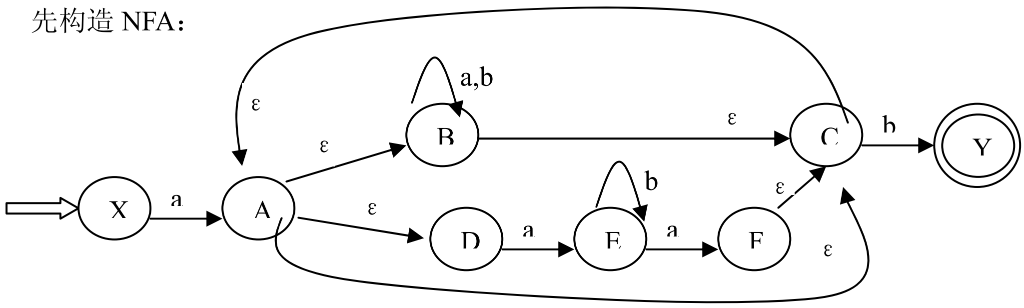
	ε	0	1
X	X		
T_0=X			A
A	ABFL		
T_1= ABFL		Y	CG
Y	Y		
CG	CGJ		
T_2= Y			
T_3= CGJ		DH	K
DH	DH		
K	ABFKL		
T_4= DH			EI
EI	ABEFIL		
T_5= ABFKL		Y	CG
T_6= ABEFIL		EJY	CG
EJY	ABEFGJLY		
T_7= ABEFGJLY		EHY	CGK
EHY	ABEFHLY		
CGK	ABCFGJKL		
T_8= ABEFHLY		EY	CGI
EY	ABEFLY		
CGI	CGJI		
T_9= ABCFGJKL		DHY	CGK
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：
先构造 NFA：

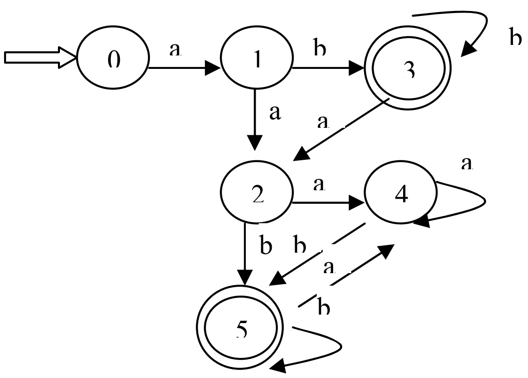


用子集法将 NFA 确定化

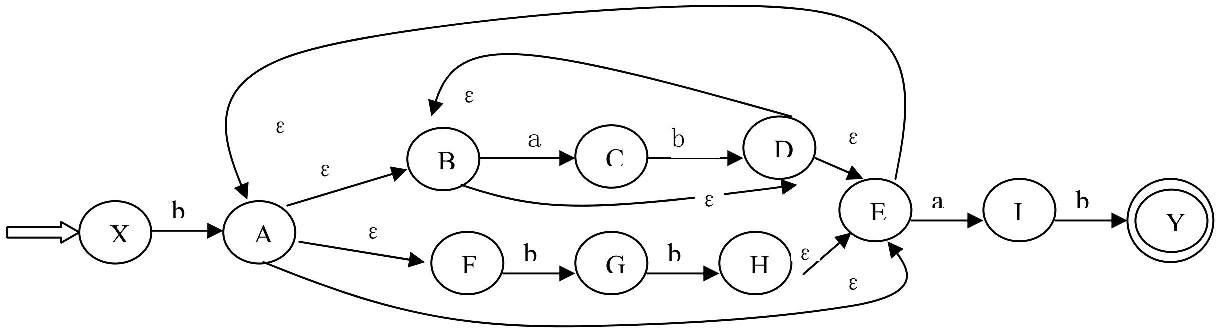
	ε	a	b
X	X		
T_0=X		A	
A	ABCD		
T_1=ABCD		BE	BY
BE	ABCDE		
BY	ABCDY		
T_2=ABCDE		BEF	BEY
BEF	ABCDEF		
BEY	ABCDEY		
T_3=ABCDY		BE	BY
T_4=ABCDEF		BEF	BEY
T_5=ABCDEY		BEF	BEY

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	4	5



(4) 先构造 NFA：



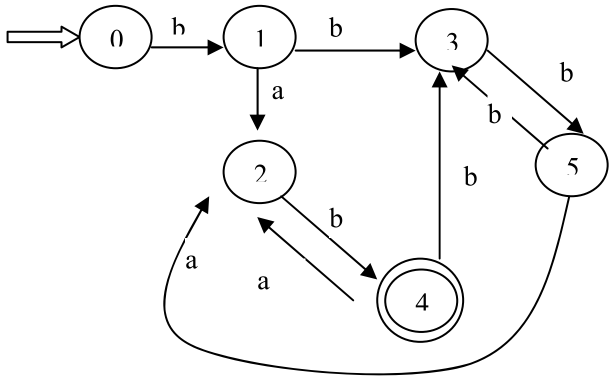
用子集法将 NFA 确定化：

	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：


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

第 2 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

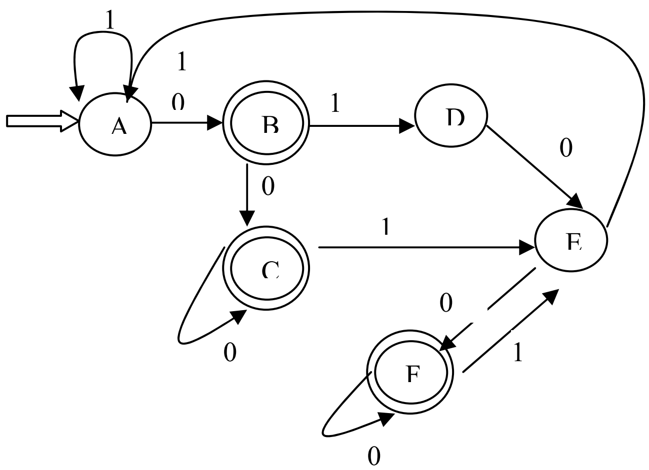
用子集法将 NFA 确定化：

	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。

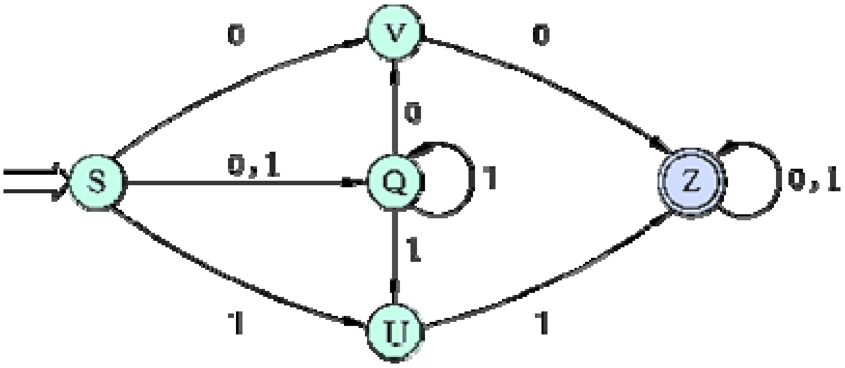
	0	1
A	B	A
B	C	D
C	C	E
D	E	
E	F	A
F	F	E

DFA 的状态图：



第 3 题

将下图确定化：



答案：

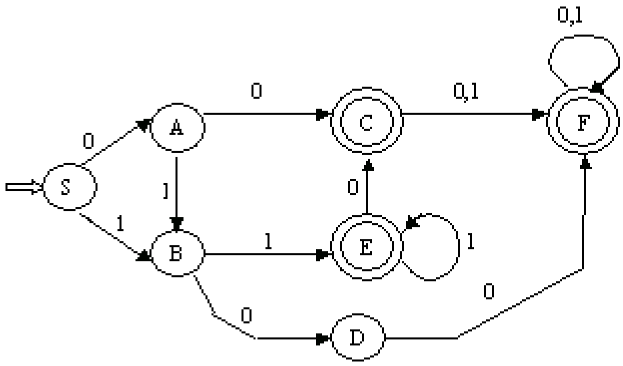
用子集法将 NFA 确定化：

.	0	1
S	VQ	QU
VQ	VZ	QU
QU	V	QUZ
VZ	Z	Z
V	Z	.
QUZ	VZ	QUZ
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。

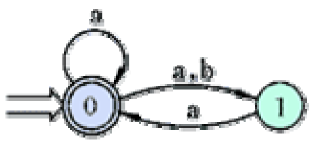
.	0	1
S	A	B
A	C	B
B	D	E
C	F	F
D	F	.
E	C	E
F	F	F

DFA 的状态图：

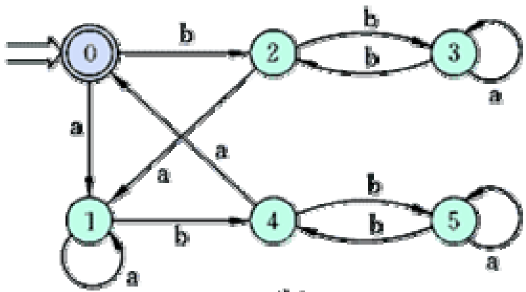


第 4 题

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}

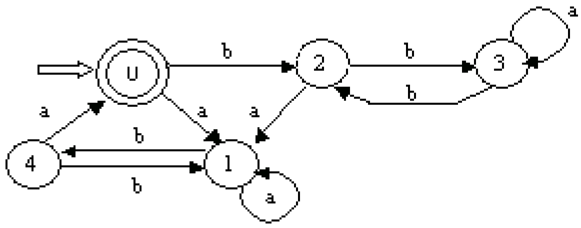
{1, 5} a ⊂{1, 5}

{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：

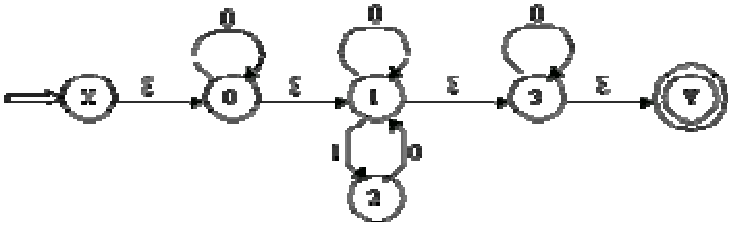


第 5 题

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



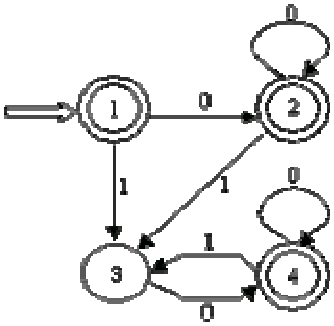
用子集法确定化：

I	I0	I1
{X,0,1,3,Y}	{0,1,3,Y}	{2}
{0,1,3,Y}	{0,1,3,Y}	{2}
{2}	{1,3,Y}	
{1,3,Y}	{1,3,Y}	{2}

重新命名状态集：

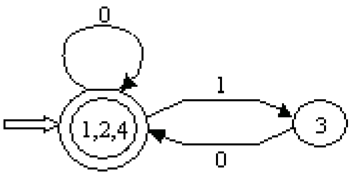
S	0	1
1	2	3
2	2	3
3	4	
4	4	3

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 6 题

设无符号数的正规式为 θ：

θ =dd*|dd*.dd*|.dd*|dd*10(s|ε)dd*

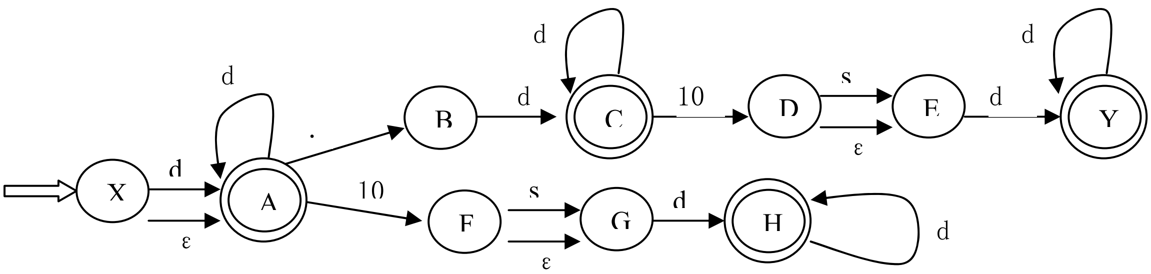
|10(s|ε)dd*|.dd*10(s|ε)dd*

|dd*.dd*10(s|ε)dd*

化简 θ，画出 θ 的 DFA，其中 d={0,1,2,…,9}，s={＋，－}

答案：

先构造 NFA：



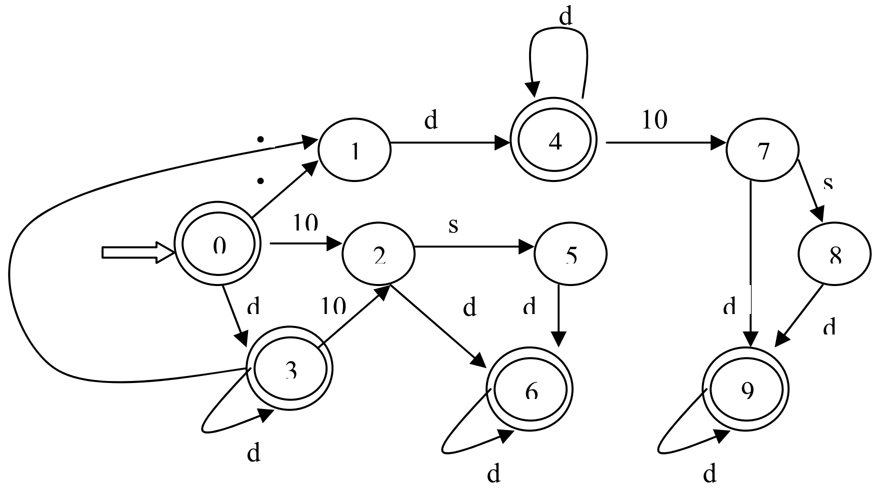
用子集法将 NFA 确定化：

	ε	•	s	10	d
X	XA				
T_0=XA		B		F	A
B	B				
F	FG				
A	A				
T_1=B					C
C	C				
T_2=FG			G		H
G	G				
H	H				
T_3=A		B		F	A
T_4=C				D	C
D	DE				
T_5=G					H
T_6=H					H
T_7=DE			E		Y
E	E				
Y	Y				
T_8=E					Y
T_9=Y					Y

将 XA、B、FG、A、C、G、H、DE、E、Y 重新命名，分别用 0、1、2、3、4、5、6、7、8、9 表示。终态有 0、3、4、6、9。

	•	s	10	d
0	1		2	3
1				4
2		5		6
3	1		2	3
4			7	4
5				6
6				6
7		8		9
8				9
9				9

DFA 的状态图：

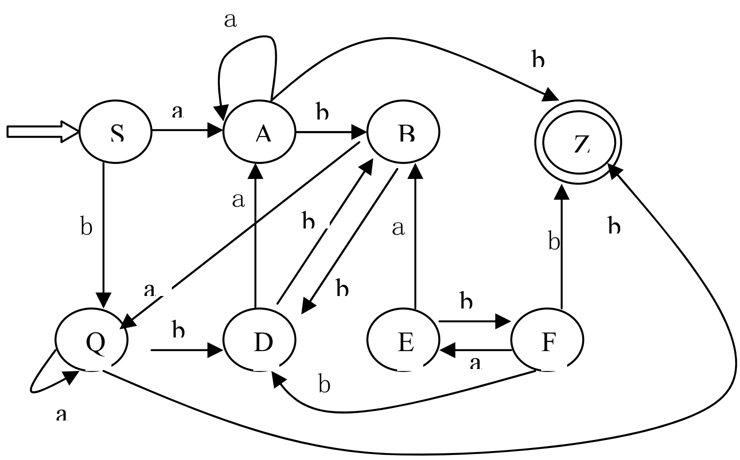


第 7 题

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：



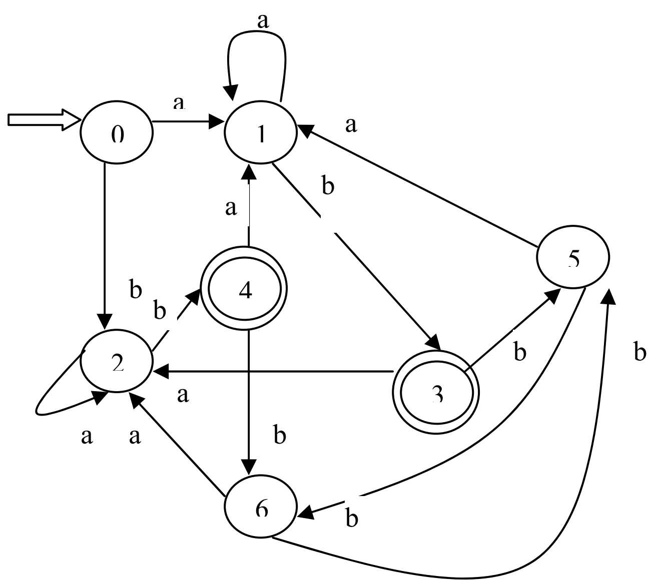
用子集法将 NFA 确定化：

	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：



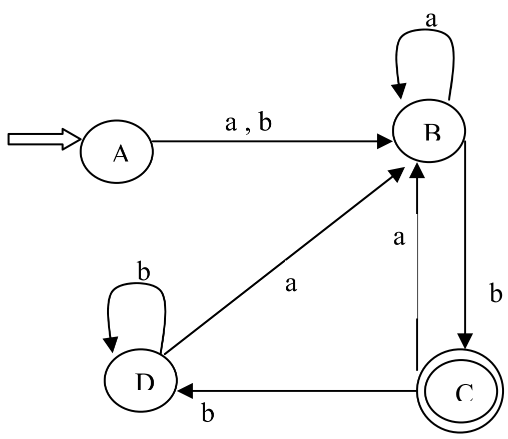
令P_0＝（｛0,1,2,5,6｝,｛3,4｝）用b进行分割：

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→0A|1B

A→1S|1

B→0S|0

答案：

解方程组 S 的解：

S=0A|1B

A=1S|1

B=0S|0

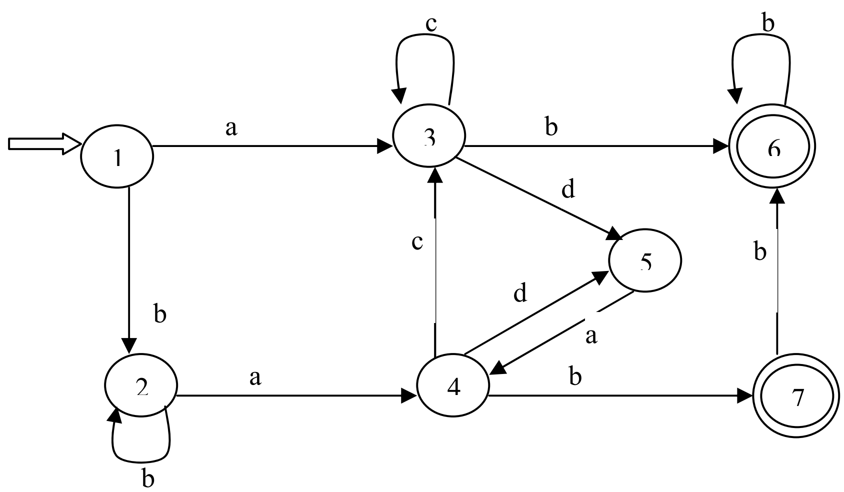
将 A、B 产生式的右部代入 S 中

S=01S|01|10S|10=（01|10）S|（01|10）

所以：S= (01|10)*(01|10)

第 9 题

将下图的 **DFA** 最小化，并用正规式描述它所识别的语言。



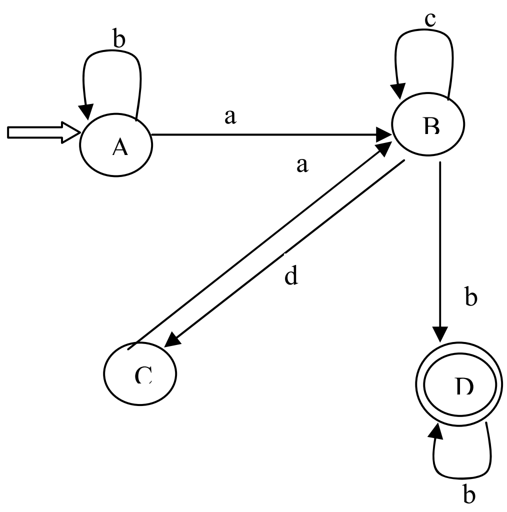
答案：

令P_0＝（｛1,2,3,4,5｝,｛6,7｝）用b进行分割：

P_1＝（｛1,2｝,｛3,4｝,｛5｝,｛6,7｝）再用a、b、c、d进行分割，仍不变。

再令｛1,2｝为 A，｛3，4｝为 B，｛5｝为 C，｛6，7｝为 D。

最小化为：



$r=b^*a(c|da)^*bb^*=b^*a(c|da)^*b^+$

附加题

问题 1：

为下边所描述的串写正规式，字母表是 {a,b}.

a) 以 ab 结尾的所有串
b) 包含偶数个 b 但不含 a 的所有串
c) 包含偶数个 b 且含任意数目 a 的所有串
d) 只包含一个 a 的所有串
e) 包含 ab 子串的所有串
f) 不包含 ab 子串的所有串

答案：

注意 正规式不唯一
a) (a|b)*ab
b) (bb)*
c) (a*ba*ba*)*
d) b*ab*
e) (a|b)*ab(a|b)*
f) b*a*

问题2：

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

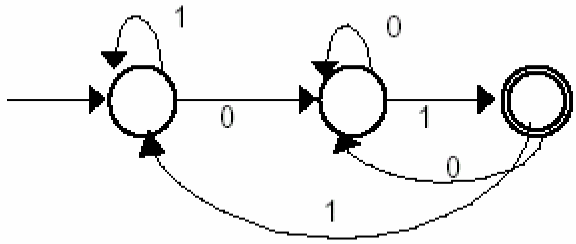
问题3：

构造有穷自动机.

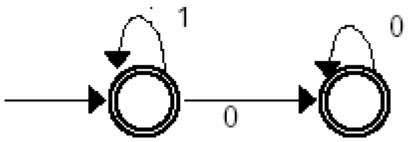
a) 构造一个 DFA，接受字母表 {0, 1}上的以 01 结尾的所有串
b) 构造一个DFA，接受字母表 {0, 1}上的不包含01 子串的所有串.
c) 构造一个NFA，接受字母表 {x,y}上的正规式x(x|y)*x描述的集合
d) 构造一个NFA，接受字母表 {a, b}上的正规式(ab|a)*b+描述的集合并将其转换为等价的DFA.以及最小状态DFA

答案：

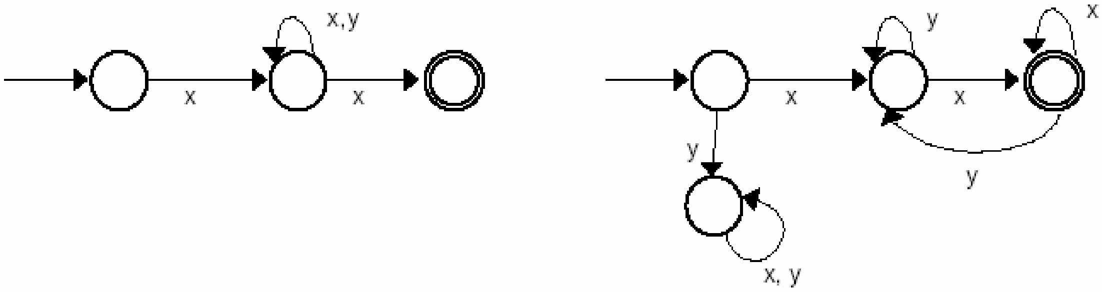
a)



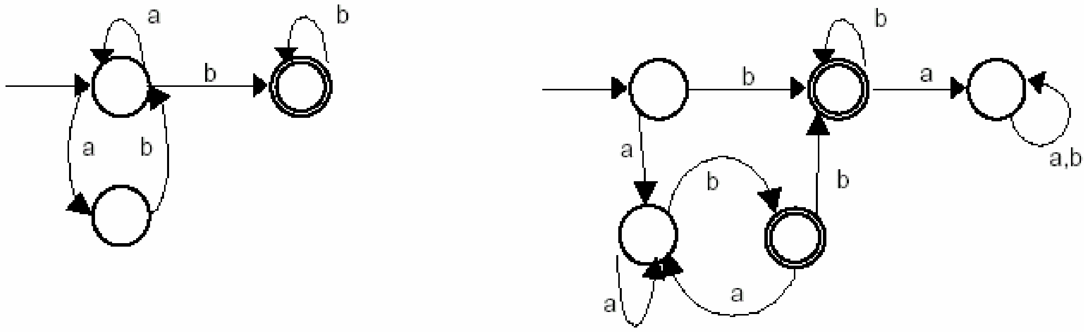
b)



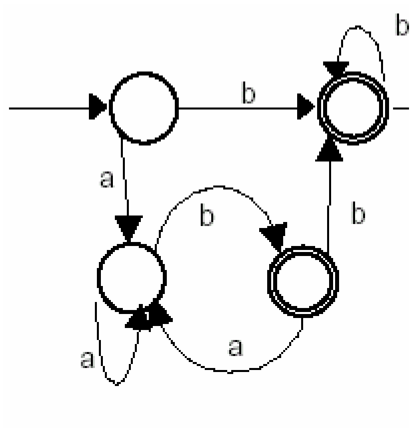
c) NFA: DFA



d) NFA: DFA

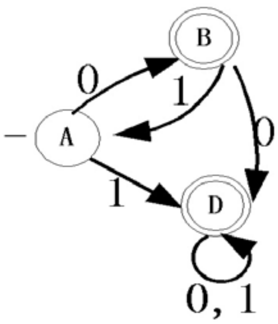


最小化的 DFA



问题 4：

设有如图所示状态转换图，求其对应的正规表达式。



可通过消结法得出正规式
R=(01)*((00|1)(0|1)*|0)
也可通过转换为正则文法，解方程得到正规式。

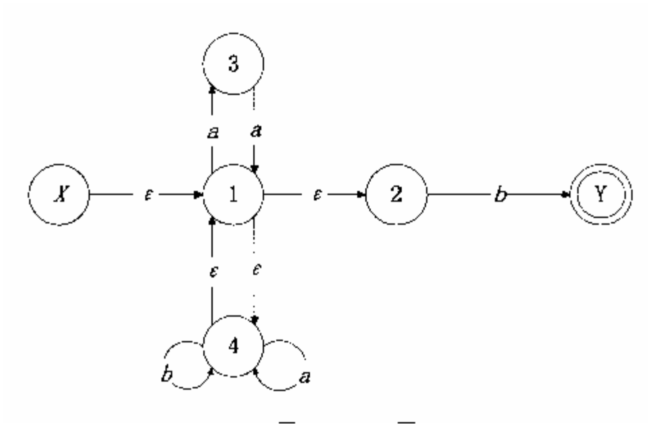
问题 5：

已知正规式：
(1)((a|b)*|aa)*b;
(2)(a|b)*b.
试用有限自动机的等价性证明正规式(1)和(2)是等价的，并给出相应的正规文法。

分析：

基本思路是对两个正规式，分别经过确定化、最小化、化简为两个最小 DFA，如这两个最小 DFA 一样，也就证明了这两个正规式是等价的。

答案：



状态转换表 1

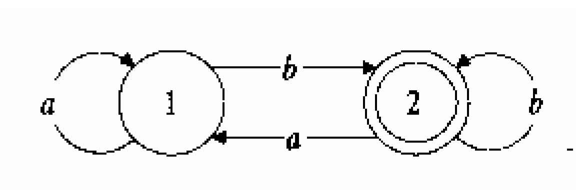
	a	b
X124	1234	124Y
1234	1234	124Y
124Y	1234	124Y

状态转换表 2

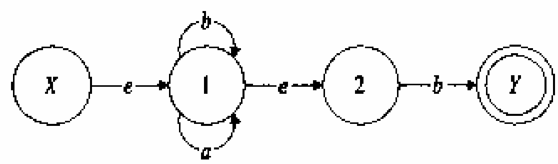
	a	B
1	2	3
2	2	3
3	2	3

由于 2 与 3 完全一样，将两者合并，即见下表

	a	b
1	2	3
	2	3



而对正规式(2)可画 NFA 图，如图所示。

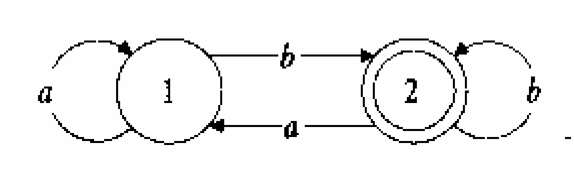


	a	b
X12	12	12Y
12	12	12Y
12Y	12	12Y

可化简得下表

	a	b
1	2	3
2	2	3

得 DFA 图



两图完全一样，故两个自动机完全一样，所以两个正规文法等价。
对相应正规文法,令 A 对应 1，B 对应 2
故为：

$A \to aA|bB|b$

$B \to aA|bB|b$

即为 $S \to aS|bS|B$,此即为所求正规文法。

问题 6:

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:

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

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

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

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

所以，一个可能的正规文法为 G[S]:

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

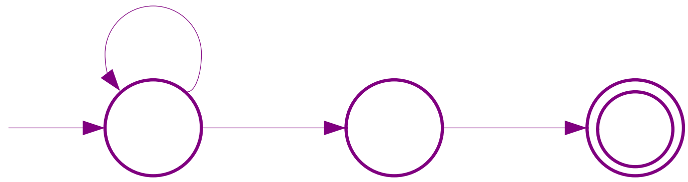
或表示为：

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

（适当等价变换也可以，但要作说明，即要有步骤）

问题 7：

考虑下图所示的 **NFA N**，构造可以生成语言 **L(N)** 的一个正规文法。



答案：

G[P]:

P → 0 P | 1 P | 1 Q

Q → 0 R | 1 R

R → ε

0, 1

问题 8：

考虑如下文法 **G[S]**：

Start

p

1

S → 0S | 1S | 1A
A→ 0B | 1B
B → ε

a) 试构造语言为 **L(G)** 的一个正规表达式。

b) 试构造语言为 **L(G)** 的一个有限自动机。

答案：

a)

由 A→ 0B ， B → ε 得 A→ 0 ;

由 A→ 1B ， B → ε 得 A→ 1 ;

由 A→ 0，A→ 1 得 A→ 0 | 1 ;

由 S→ 1A，A→ 0 | 1 得 S→ 1(0 | 1) ;

由 S→ 1A，A→ 0 | 1 得 S→ 1(0 | 1) ;

由 S→ 0S，S→ 1(0 | 1) 得 S→ 0*1(0 | 1) ;

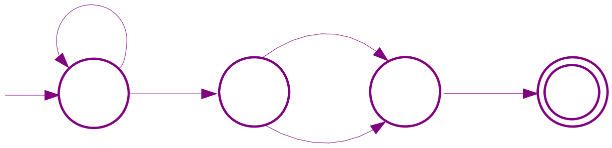
由 S→ 1S，S→ 1(0 | 1) 得 S→ 1*1(0 | 1) ;

由 S→ 0*1(0 | 1)，S→ 1*1(0 | 1) 得 S→ 0*1(0 | 1) | 1*1(0 | 1) ;

所以，一个可能的正规表达式为：

0*1(0 | 1) | 1*1(0 | 1)

b)



0，1

Start S 1 A