

编译原理

Compiler Construction Principles



朱 青

信息学院计算机系，
中国人民大学，

zqruc2012@aliyun.com



复习：词法分析

Lexical Analysis

第2章:词法分析 (Lexical Analysis)

⌘ 2.1 词法分析程序的功能

⌘ 2.2 词法分析器的设计

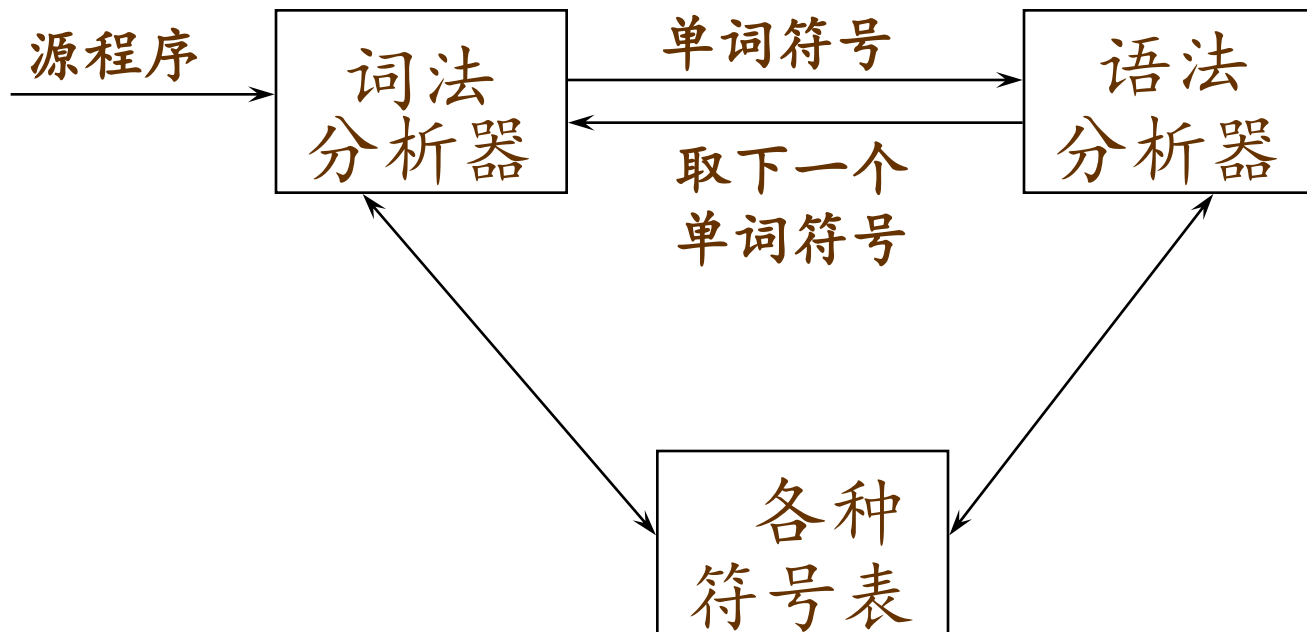
⌘ 2.3 正规表达式 (Regular Expression)

⌘ 2.4 有限自动机

⌘ 2.5 词法分析器的自动生成

第2章:词法分析实验

词法分析是编译的基础。



第2章:词法分析复习

词法分析总结与习题:

基本概念

1. 编译过程 编译程序
2. 解释程序 汇编程序
3. 翻译程序
4. 编译程序结构(各阶段输入,输出)

5.状态转换图,转换矩阵.

6.正规式,正规集,正规文法.

7.DFA,NFA,LEX.

实验性题目:

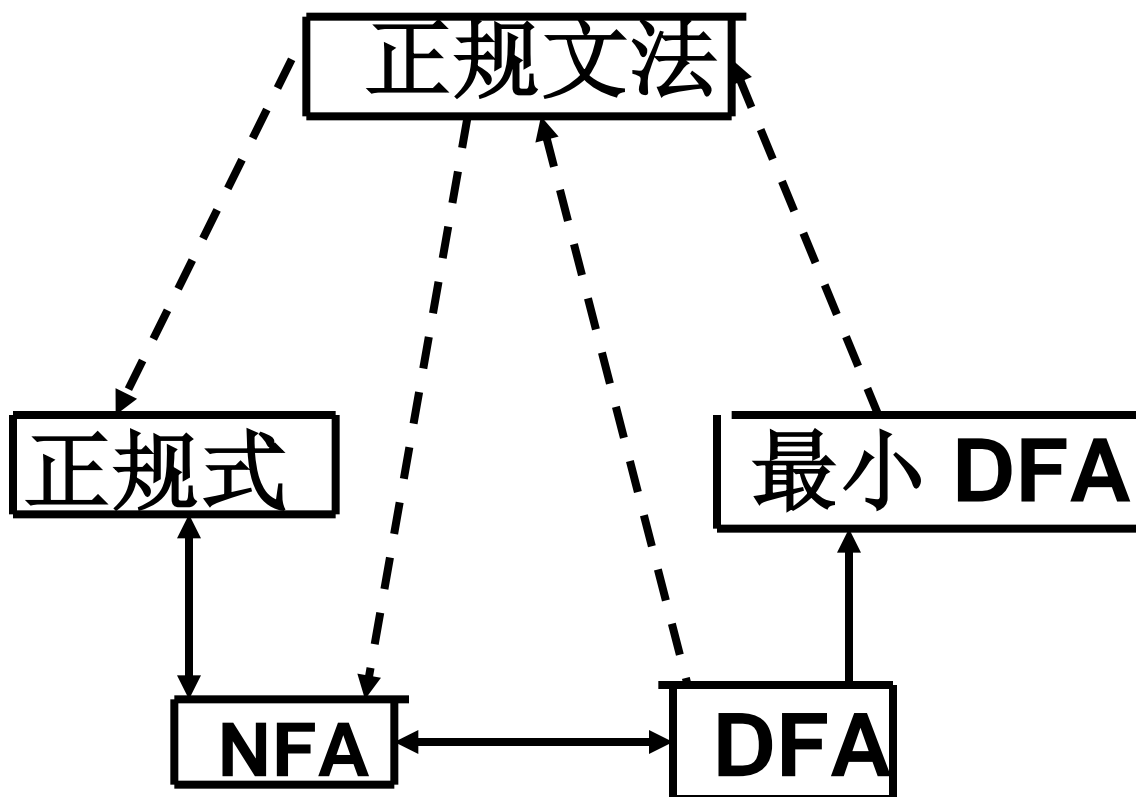
1 状态转换图的实现.

2 词法分析器的自动生成.

理论性题目:

1. 状态转换图.
2. 正规式与正规集.
3. 确定有限自动机.(DFA)
4. 非确定有限自动机.(NFA)
5. 正规文法.

正规集各种描述工具间的转换:



一：根据要求画出状态转换图：

例题：

(1)识别标识符的状态转换图.

(2)识别**FORTRAN**实型常数的状态转换图.(P41)

(3)识别某个简单语言的所有单词的状态转换图.

二. 试描述下列各正规表达式所表示的语言:

例题: (1) $(00|01|10|11)^*$

(2) $(0|1)(0|1)(0|1)^*$

解:(1) 任何长度为偶数或长度为0的0,1符号串.

(2)任何长度大于等于2的0,1符号串.

三.试证明两个正规表达式等价:

例题:(P63) 6.

令A,B,C是任意正规式,证明下列关系成立:

$$(5) \quad (A|B)^* = (A^*B^*)^* = (A^*|B^*)^*$$

$$(6) \quad A = b|aA \quad \text{当且仅当} \quad A = a^*b$$

证明:(5)

先证: $(A|B)^* = (A^*B^*)^*$

因为 $L(A) \subset L(A)^*L(B)^*$
 $L(B) \subset L(A)^*L(B)^*$

所以 $L(A) \cup L(B) \subseteq L(A)^*L(B)^*$

于是 $(L(A) \cup L(B))^* \subseteq (L(A)^*L(B)^*)^*$

又因为 $L(A) \subseteq L(A) \cup L(B)$

所以 $L(A)^* \subseteq (L(A) \cup L(B))^*$

同样 $L(B)^* \subseteq (L(A) \cup L(B))^*$

所以

$L(A)^* L(B)^* \subseteq (L(A) \cup L(B))^* (L(A) \cup L(B))^*$

所以 $L(A)^* L(B)^* \subseteq ((L(A) \cup L(B))^*)^2$

因此 $(L(A)^* L(B)^*)^* \subseteq ((L(A) \cup L(B))^*)^*$

由6.2 $(L(A)^* L(B)^*)^* \subseteq (L(A) \cup L(B))^*$

于是 $(L(A) \cup L(B))^* = (L(A)^* L(B)^*)^*$

即有: $(A|B)^* = (A^*B^*)^*$

再证: $(A|B)^* = (A^*|B^*)^*$

我们换一种方法证明:

因为 $(A^*|B^*)^* = ((A^*)|(B^*))^*$

由刚刚的证明得:

$$\begin{aligned} ((A)^*|(B)^*)^* &= ((A^*)(B^*))^* \\ &= (A^*B^*)^* \end{aligned}$$

则: $(A^*|B^*)^* = (A^*B^*)^*$

于是得证:

$$(A|B)^* = (A^*B^*)^* = (A^*|B^*)^*$$



证明:(6) $A=b|aA$ 当且仅当 $A=a^*b$

首先: 由 $A=a^*b$ 推证 $A=b|aA$

$$\begin{aligned}\text{因为 } a^*b &= (a^0 | a^1 | a^2 | \dots) b \\ &= (a^0b | a^1b | a^2b | \dots) \\ &= (\varepsilon b | a^1b | a^2b | \dots) \\ &= b|a(b | a^1b | a^2b | \dots) \\ &= b|a((\varepsilon | a^1 | a^2 | \dots) b) \\ &= b|a(a^*b) = b|aA\end{aligned}$$

故有: $A=b|aA$

然后 由 $A=b|aA$ 推证 $A=a^*b$

因为 $A=b|aA$

亦即 $A=b$ 或有 $A=aA$

则正规式 A 表示的字集 Σ^* 由下列字组成

$b, ab, aab, aaab, \dots, aa \dots ab$

因为 $b \in \Sigma^*$

$$\begin{aligned}\text{又由 } A &= aA = a(aA) = a(a(aA)) \\ &= a(a \dots (aA) \dots)\end{aligned}$$

所以 A 表示的字字首含有若干个 a .

而由 $A=b$ 知 A 表示的字以 b 为尾.

于是, A 表示的字就是以若干个 a 开头, 而以字符 b 为尾的字符串.

亦即 $a^0b, ab, a^2b, a^3b, \dots$

故得 $A = a^*b$ |

四.对于下列各语言分别写出他们的
正规表达式:

例如:(1)任何以**abb**为结尾的**a,b**符号串.

(2) $\{a,aa,aaa,aaaa,\dots\}$

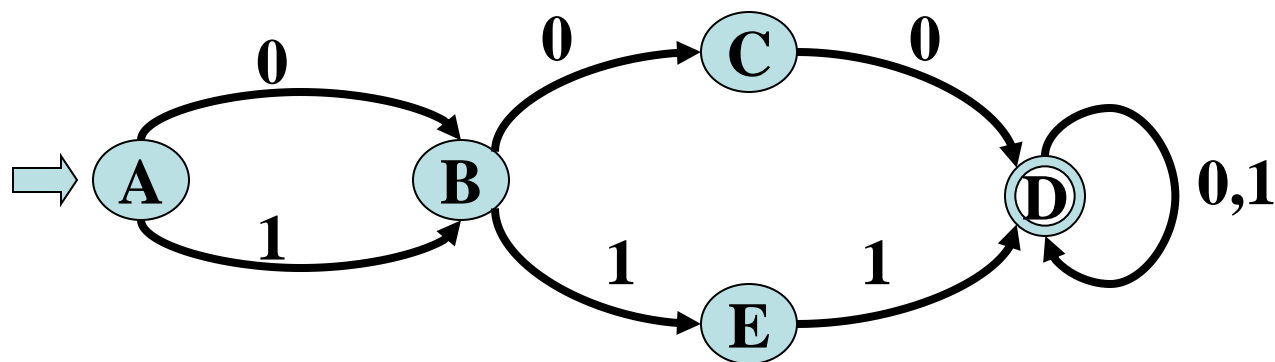
解: (1) $(a|b)^*abb$

(2) aa^*

五.给出接受下列在字母表 $\{0,1\}$ 上的语言的DFA:

(1) 接受的语言 $L(M)$ 是由 $\{0,1\}^*$ 中所有含有二个相继为0或二个相继为1的符号串组成的集合,其二个相继符号串最早在第二三位开始出现.(例:1001011)

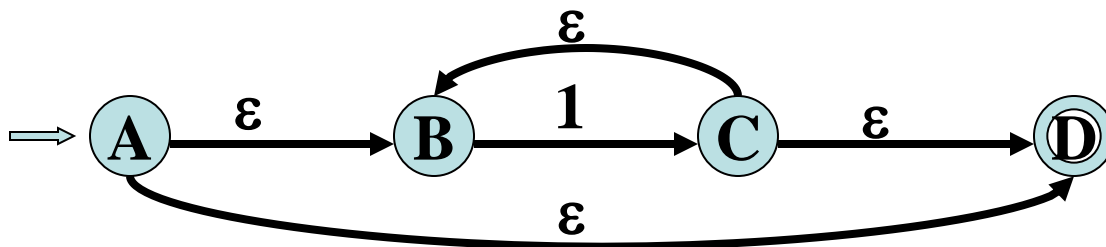
解：其正规式： $(0|1)(00|11)(0|1)^*$



六. 构造等价于下列正规表达式的有限自动机:

(1) $r1=1^*$

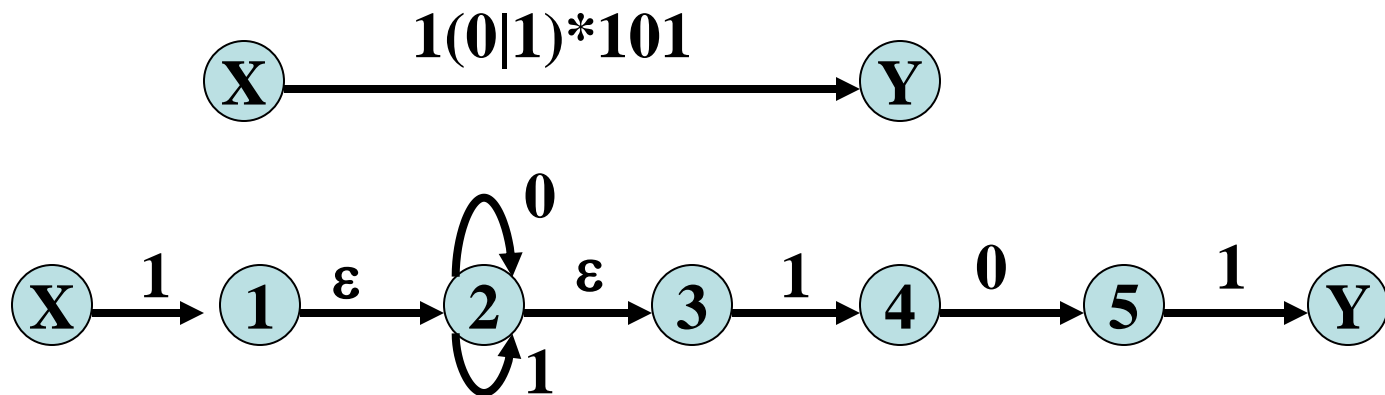
解:(1)



七. 对于下列正规表达式构造 非确定有限自动机:

(1) $1(0|1)^*101$

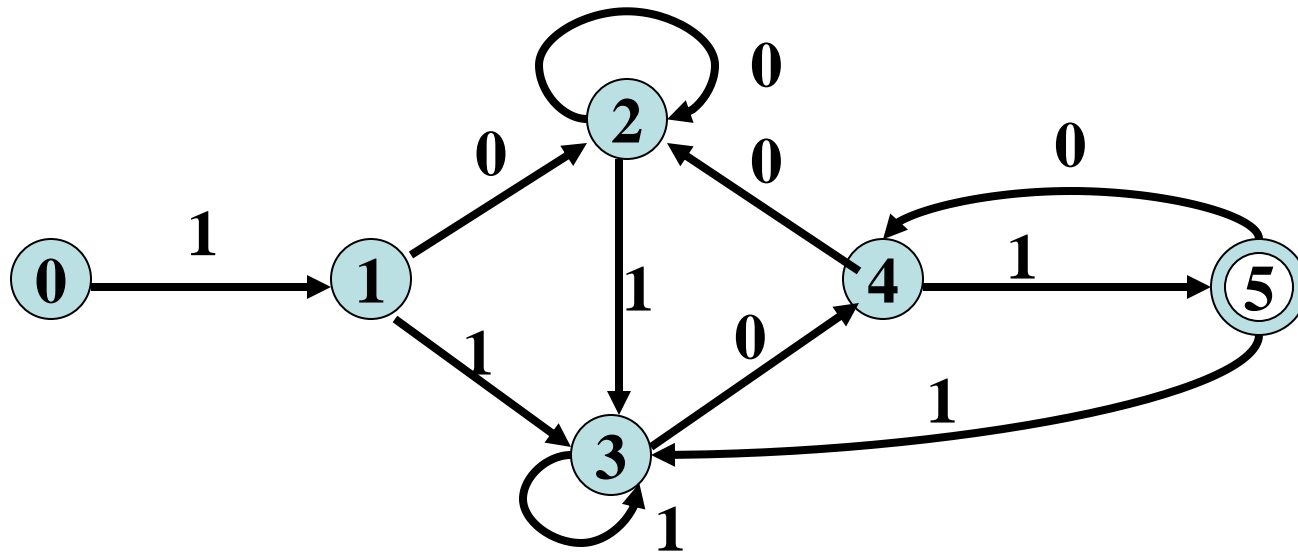
解:



八. 将非确定有限自动机确定化：（前例）

I	I_0	I_1	S	0	1
{X}		{1,2,3}	0		1
{1,2,3}	{2,3}	{2,3,4}	1	2	3
{2,3}	{2,3}	{2,3,4}	2	2	3
{2,3,4}	{2,3,5}	{2,3,4}	3	4	3
{2,3,5}	{2,3}	{2,3,4,Y}	4	2	5
{2,3,4,Y}	{2,3,5}	{2,3,4}	5	4	3

DFA M



九.将确定有限自动机最小化: (前例)

解:

首先分成: $\{0,1,2,3,4\}$ $\{5\}$

$$\{0,1,2,3,4\}_1 = \{1,3,5\}$$

分成: $\{1,2\}$, $\{0,3,4\}$, $\{5\}$

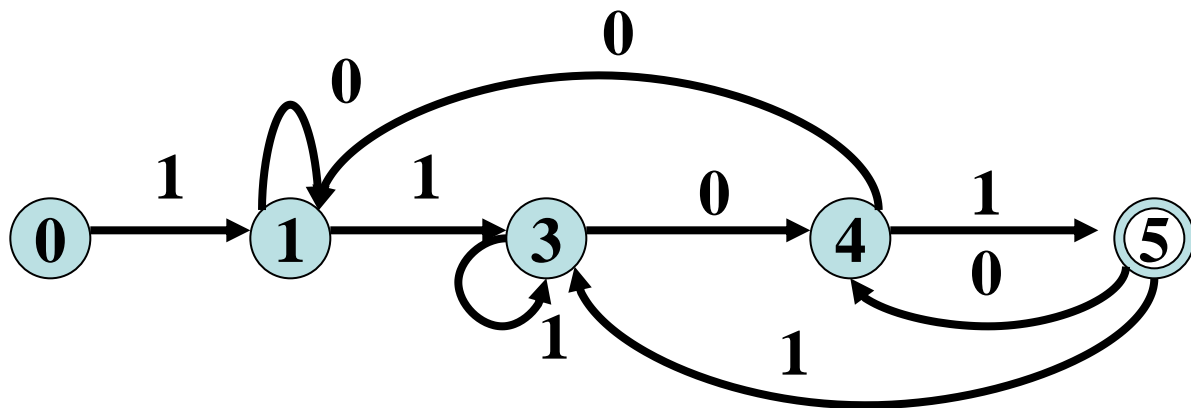
$$\{1,2\}_0 = \{2\} \quad \{1,2\}_1 = \{3\} \quad \text{不再分}$$

$$\{0,3,4\}_0 = \{4,2\} \text{ 分成: } \{0,3\}, \{4\}$$

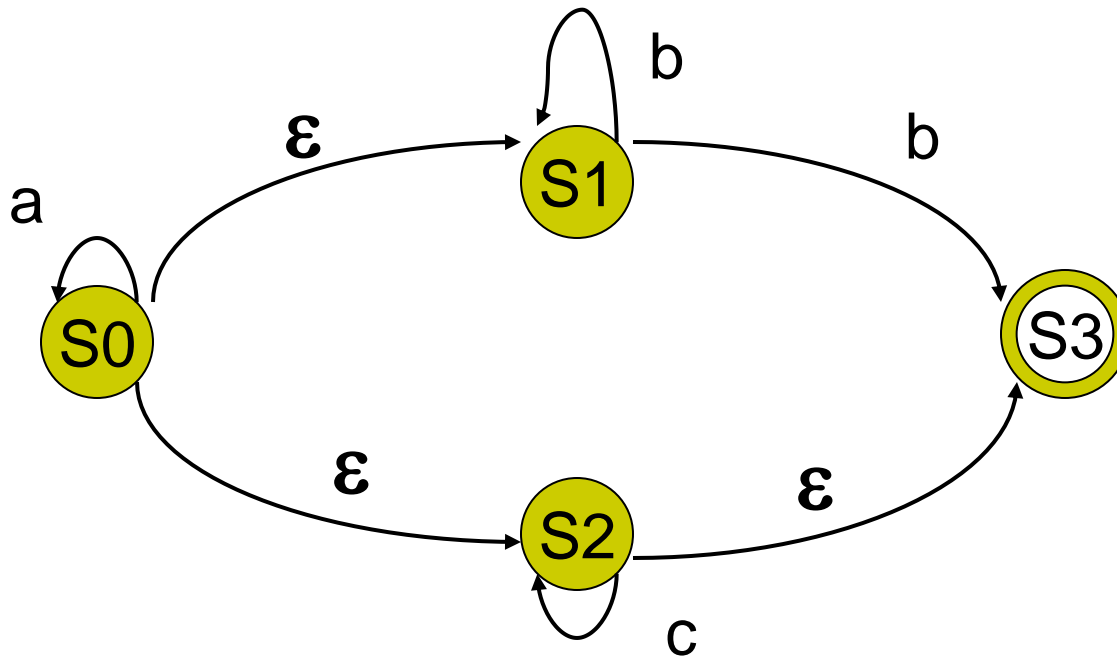
分成: $\{1,2\}$, $\{0,3\}$, $\{4\}$, $\{5\}$

$\{0,3\}_1 = \{1,3\}$ 分成: $\{0\}, \{3\}$.

分成: $\{0\}, \{1,2\}, \{3\}, \{4\}, \{5\}$.



例题. 将 ϵ -NFA确定化:



解:

$$1) \ \varepsilon\text{-closure}(S0) = \{S0, S1, S2, S3\}$$

记作: $q0 \dashrightarrow \{S0, S1, S2, S3\}$

$$2) \ f'(q0, a) = \varepsilon\text{-closure}(S0) = q0$$

$$\begin{aligned} f'(q0, b) &= \varepsilon\text{-closure}(\{S1, S3\}) \\ &= \{S1, S3\} \end{aligned}$$

记作: $q1 \dashrightarrow \{S1, S3\}$

$$\begin{aligned} f'(q_0, c) &= \varepsilon\text{-closure}(\{S_2\}, c) \\ &= \{S_2, S_3\} \end{aligned}$$

记作: $q_2 \dashv \{S_2, S_3\}$

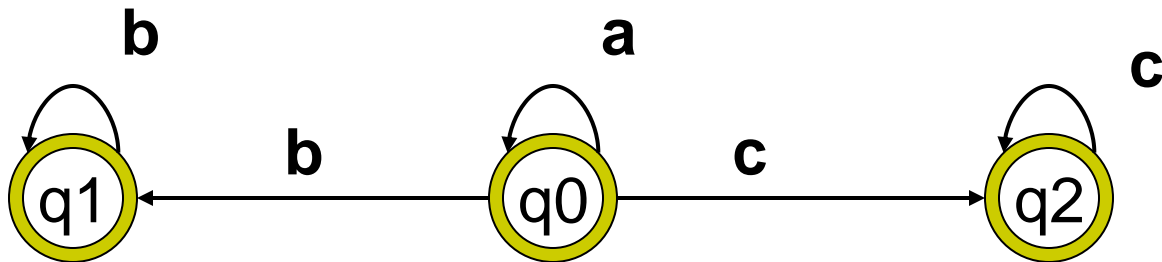
$$\begin{aligned} 3) f'(q_1, a) &= \varepsilon\text{-closure}(\{S_1, S_3\}, a) \\ &= \varepsilon\text{-closure}(\Phi) = \Phi \end{aligned}$$

$$\begin{aligned} f'(q_1, b) &= \varepsilon\text{-closure}(\{S_1, S_3\}, b) \\ &= \varepsilon\text{-closure}(\{S_1, S_3\}) \\ &= q_1 \end{aligned}$$

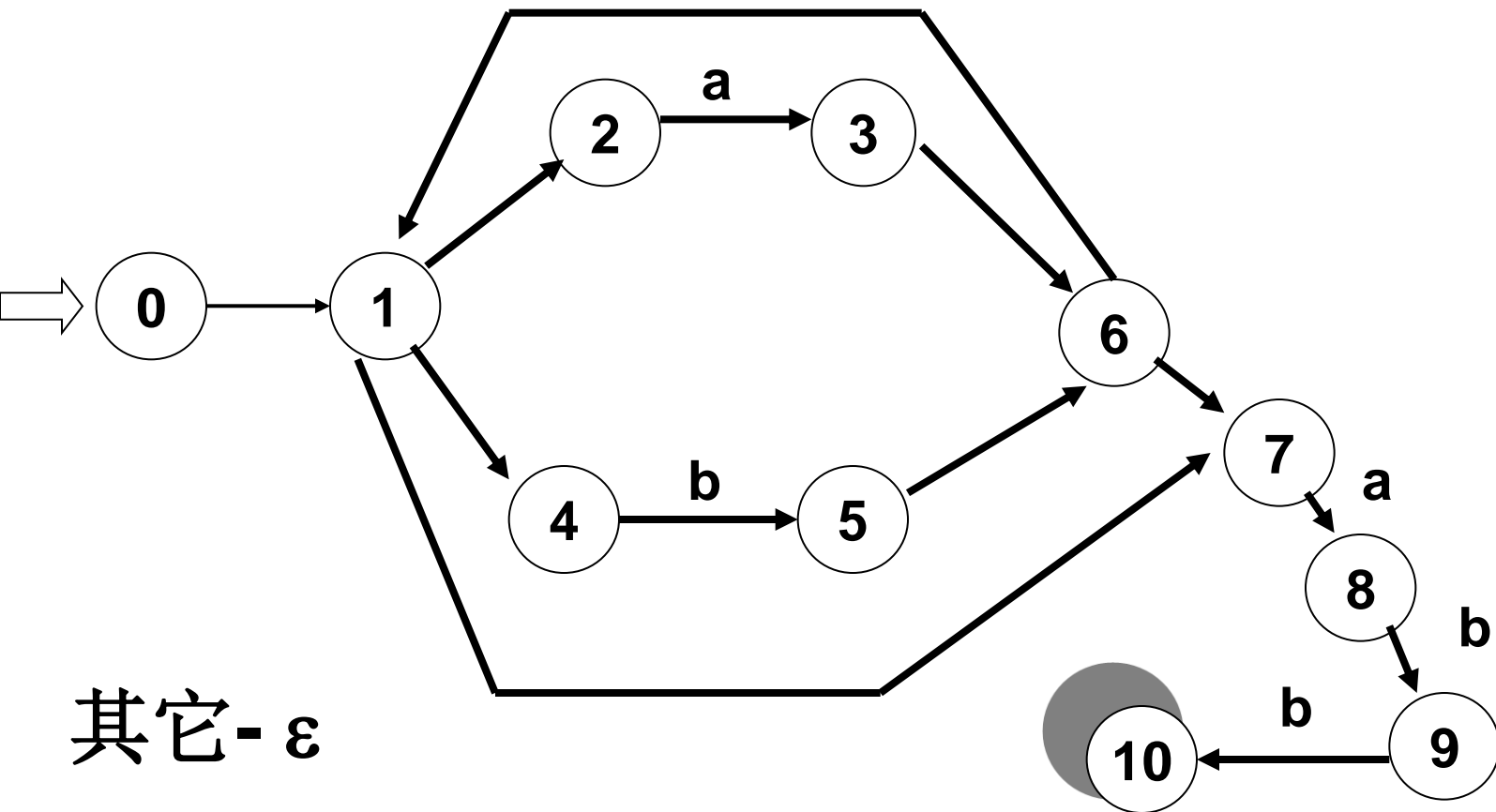
$$f'(q1,c) = \varepsilon\text{-closure}(\{S1,S3\},c) \\ = \Phi$$

$$f'(q2,a) = f'(q2,b) = \Phi$$

$$f'(q2,c) = q2$$



例: 如下一个具有 ϵ - 转移的
NFA M , 构造与之等价的 DFA M' .



解： 1) 初态

$$\varepsilon\text{-closure}(0)=\{0,1,2,4,7\}=A$$

2) 输入的字母表 $\Sigma=\{a,b\}$

3) $\varepsilon\text{-closure}(f(A,a))$

$$= \varepsilon\text{-closure}(f(\{0,1,2,4,7\},a))$$

$$= \varepsilon\text{-closure}(\{3,8\})$$

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

$$\begin{aligned}
& \varepsilon\text{-closure}(f(A,b)) \\
&= \varepsilon\text{-closure}(f(\{0,1,2,4,7\},b)) \\
&= \varepsilon\text{-closure}(\{5\}) \\
&= \{1,2,4,5,6,7\} = C
\end{aligned}$$

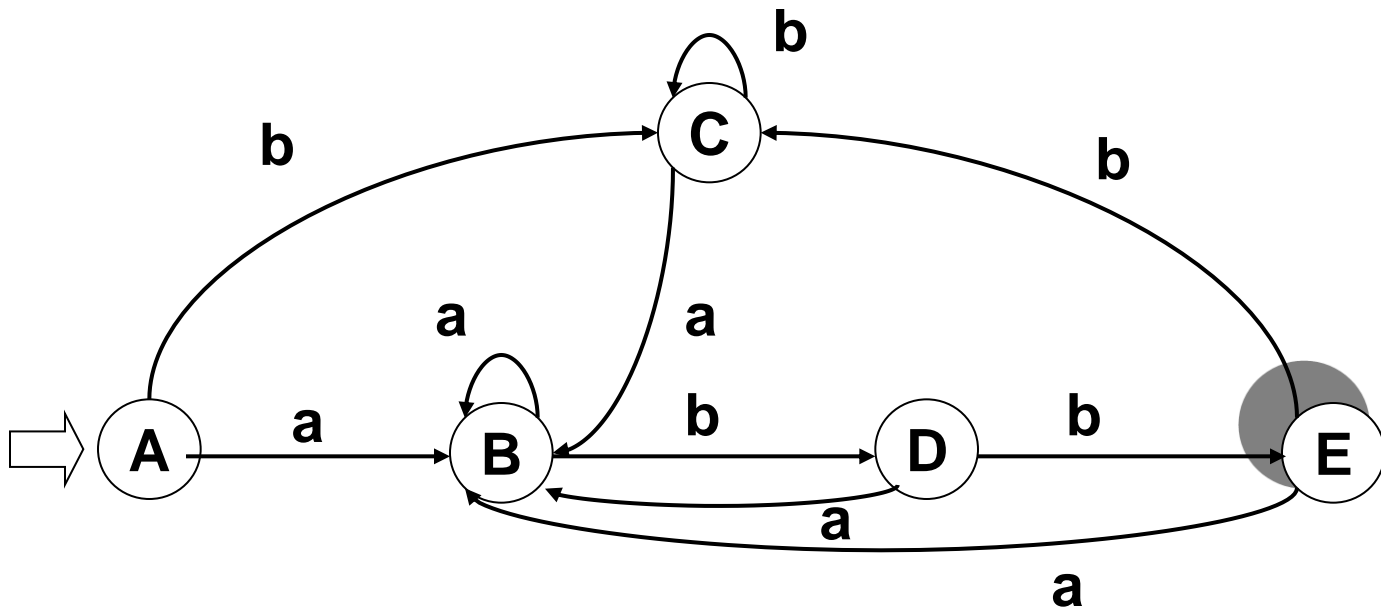
... ..

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

$$E = \{1,2,4,5,6,7,10\}$$

转换表:

状态	输入符号	
	a	b
A	B	C
B	B	D
C	B	C
D	B	E
E	B	C



练习:

1. 根据要求画出状态转换图:

画出识别不以**0**打头的正奇数的
状态转换图.

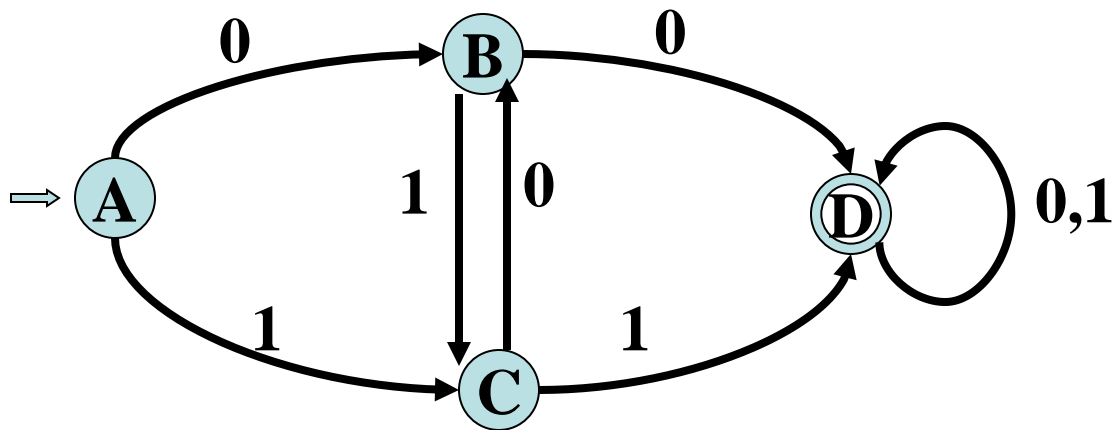
2. 对于下列语言写出其正规表达式:

(1) 用正规表达式表示所有由**a,b**组成的长度为**4**的串集.

(2) 用正规表达式表示语言 $L(M)$ 是由 $\{a,b\}^*$ 中所有二个相继为 a 或二个相继为 b 的符号串组成的集合.

3*. 写出与下面DFA等价的正规文法.

(1)



4.构造与下列正规式

(1) $r1=01^*$

(2) $r2=01^*|1$

等价的有限自动机.

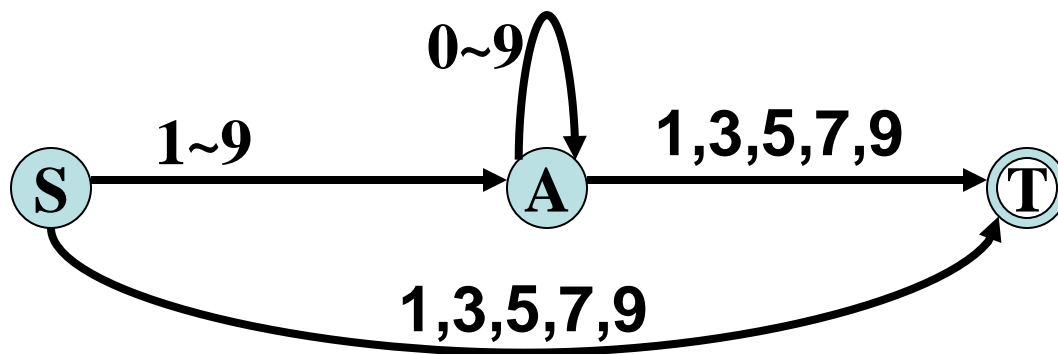
5.构造一个**DFA**,它接受 $\Sigma=\{0,1\}$,
上所有满足下列条件的字符串:
每个1都有0直接跟在右面.

**6.画出用来识别如下三个关键字的
状态转换图**

STEP STRING SWITCH

练习答案:

1.



2. (1). $(a|b) (a|b) (a|b) (a|b)$.

2(2). $(a|b)^*(aa|bb)(a|b)^*$.

3. (1)文法

$A \rightarrow 0B|1C$

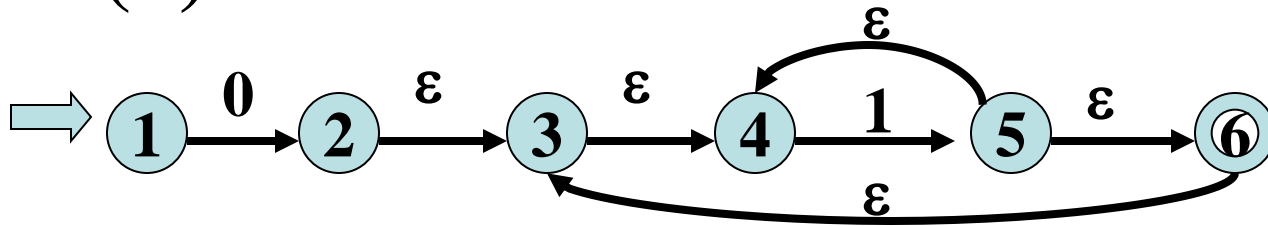
$B \rightarrow 0|0D|1C|0D$

$C \rightarrow 1|0B|1D$

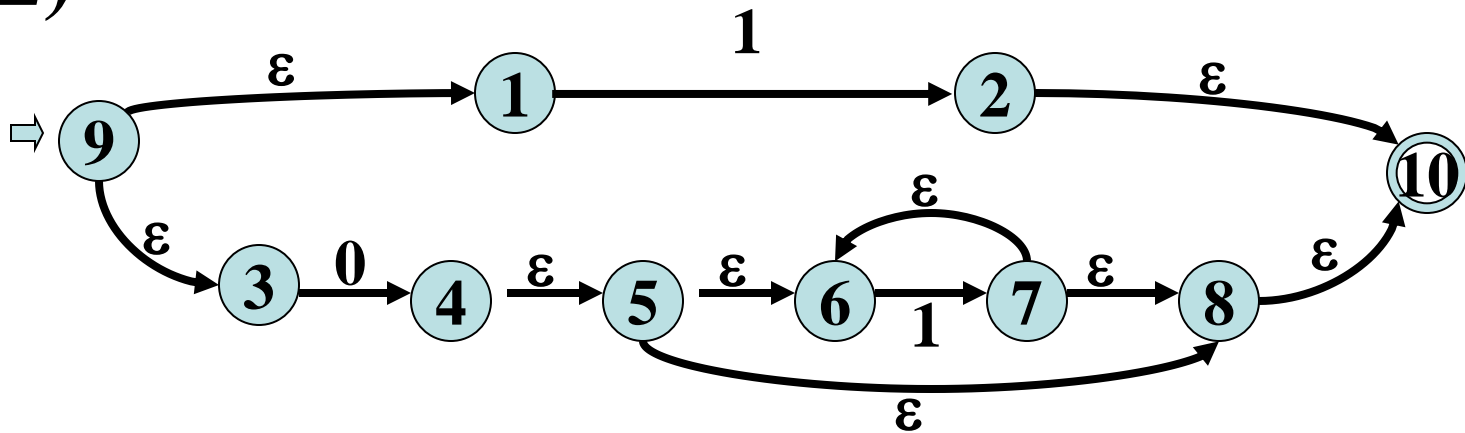
$D \rightarrow 0|1|0D|1D$

正规式: $(0|1)^*(00|11)(0|1)^*$.

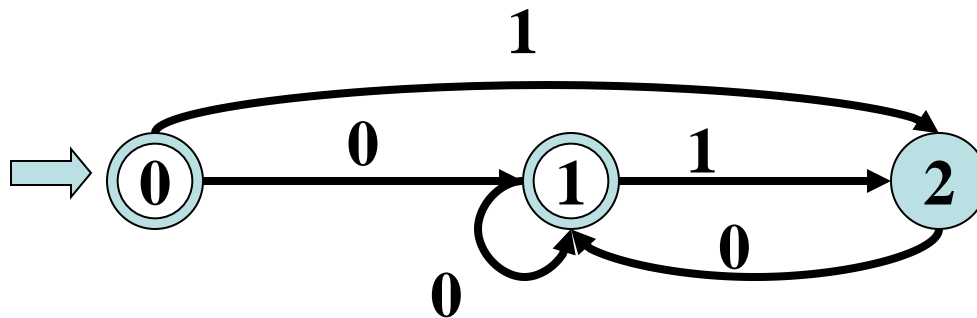
4.(1)



(2)



5. $(10|0)^*$



习题P64 12(1)

答：（确定化）转换矩阵：

	la	lb	s	a	b
{0}	{0,1}	{ 1 }	0	1	2
{0,1}	{0,1}	{1}	1	1	2
{1}	{0}		2	0	

（最小化）：

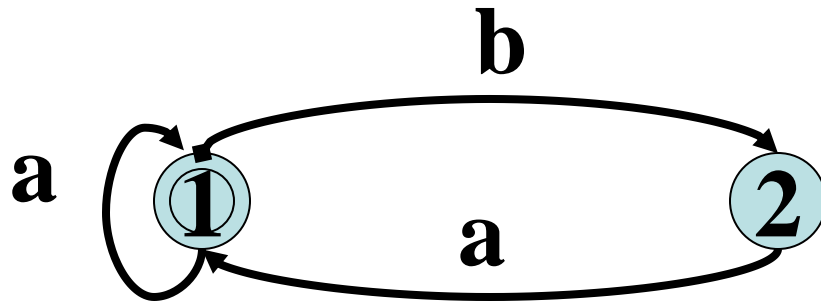
首先：终态组{0, 1}，非终态组{2}

考察 $\{0,1\}$

$\{0,1\}_a = \{1\}$ 属于 $\{0,1\}$

$\{0,1\}_b = \{2\}$ 属于 $\{2\}$

状态: $\{2\}, \{0,1\}$



习题P64 12(2)

