

# 中山大学计算机学院

## 编译原理理论

### 第一次作业

(2024学年春季学期)

课程名称:

编译原理	第一次作业	专业(方向)	计算机科学与技术
学号	21307185	姓名	张礼贤
Email	✉ <a href="mailto:zhanglx78@mail2.sysu.edu.cn">zhanglx78@mail2.sysu.edu.cn</a>	完成日期	2024.4.20

## 1. 正则表达式转换为 NFA 与子集构造

给定字母表  $\Sigma = a, b$  上的正则表达式  $(a + b)^* a(a + b)^*$ , 请完成以下题目:

### 1.1 使用 McNaughton-Yamada-Thompson 算法将上述正则表达式转换为 NFA 并绘制出来

#### 1. Thompson算法:

汤姆森算法的基本的构造规则包括:

- 将正则表达式中的每个字符转换为一个NFA片段, 每个片段包含一个状态和一个或多个转移边, 用于匹配对应的字符。
- 连接操作:** 将两个NFA片段连接起来, 形成一个新的NFA, 其中第一个片段的接受状态与第二个片段的初始状态通过  $\epsilon$ -转移相连。
- 选择操作:** 将两个NFA片段分别作为选择的两个分支, 形成一个新的NFA, 其中新的初始状态通过  $\epsilon$ -转移连接到两个原始片段的初始

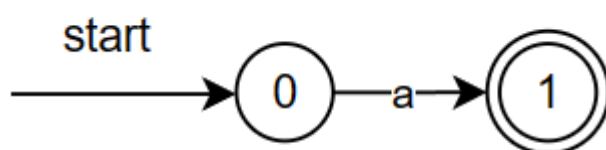
状态，而两个原始片段的接受状态分别通过  $\epsilon$ -转移连接到新的接受状态。

4. **闭包操作：** 将一个NFA片段进行闭包操作，即在其初始状态和接受状态之间添加  $\epsilon$ -转移边，形成一个新的NFA。

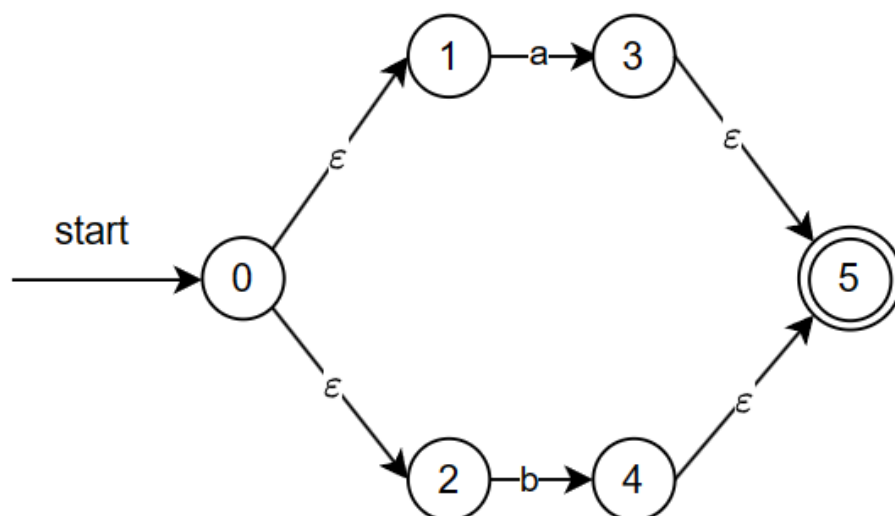
## 2. 解题过程：

通过Thompson算法，将正则表达式  $(a + b) * a(a + b)^*$  拆解成如下的几个元结构：

- $a$



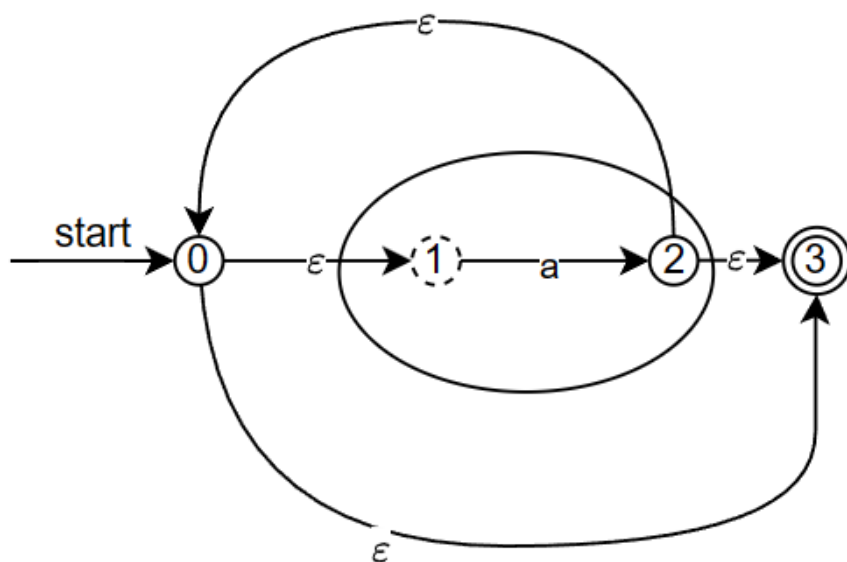
- $a + b$



- $a^*$

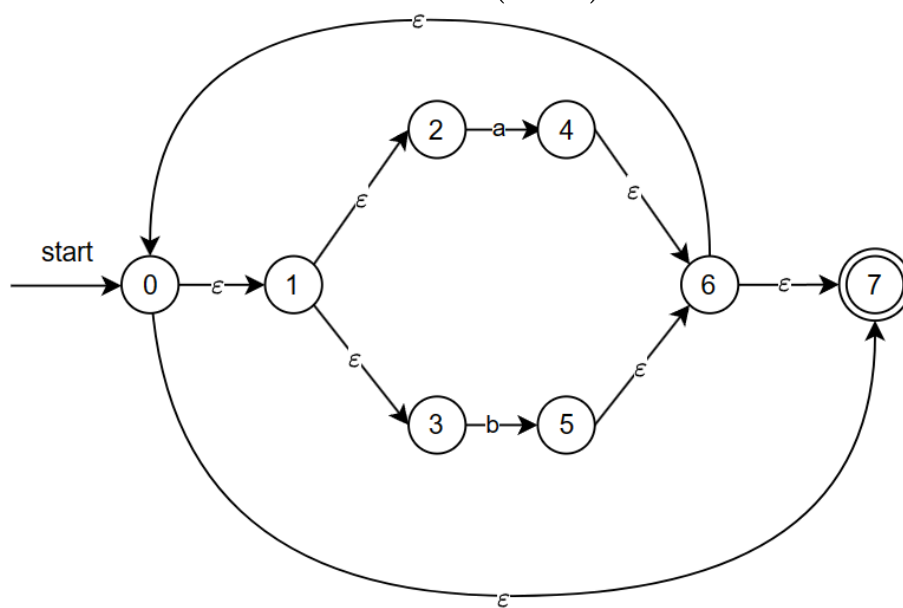
在上面的图中，椭圆圈起来的结构元可以被其他的结构元替换，1

表示初始状态，2表示接受状态



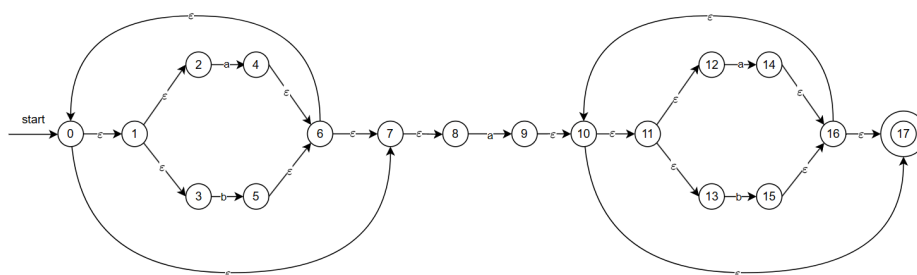
◦  $(a + b)^*$

将  $a + b$  替换为上面的结构元，得到  $(a + b)^*$  的NFA



◦  $(a + b)^* a (a + b)^*$

将上面的结构元全部连接到一起，得到最终的NFA



# 1.2 在你的 NFA 的基础上构建 DFA 转移表

NFA STATE	DFA STATE	a	b
{0, 1, 2, 3, 7, 8}	A	B	C
{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17}	B	D	E
{0, 1, 2, 3, 5, 6, 7, 8}	C	B	C
{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17}	D	D	E
{0, 1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17}	E	D	E

**转换过程：**

当新生成的状态不在DFA状态集中时， 标记为新的状态；当新生成的状态在DFA状态集中时， 直接标记为对应的状态， 并且不再放入 DFA STATE 集合中， 在这里标记为 end。

$$\epsilon\text{-closure}\{0\} = \{0, 1, 2, 3, 7, 8\}$$

$$\begin{aligned} \text{move}(\{0, 1, 2, 3, 7, 8\}, a) &= \{4, 9\} \\ \epsilon\text{-closure}\{4, 9\} &= \{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17\} \\ B &= \{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17\} \end{aligned}$$

$$\begin{aligned} \text{move}(\{0, 1, 2, 3, 7, 8\}, b) &= \{5\} \\ \epsilon\text{-closure}\{5\} &= \{0, 1, 2, 3, 5, 6, 7, 8\} \\ C &= \{0, 1, 2, 3, 5, 6, 7, 8\} \end{aligned}$$

$$\begin{aligned} \text{move}(\{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17\}, a) &= \{4, 9, 14\} \\ \epsilon\text{-closure}\{4, 9, 14\} &= \{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17\} \\ D &= \{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17\} \end{aligned}$$

$$\begin{aligned} \text{move}(\{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17\}, b) &= \{5, 15\} \\ \epsilon\text{-closure}\{5, 15\} &= \{0, 1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 15, 17\} \\ E &= \{0, 1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 15, 17\} \end{aligned}$$

$$\begin{aligned} \text{move}(\{0, 1, 2, 3, 5, 6, 7, 8\}, a) &= \{4, 9\} \\ \epsilon\text{-closure}\{4, 9\} &= \{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17\} = B \\ \text{end} \end{aligned}$$

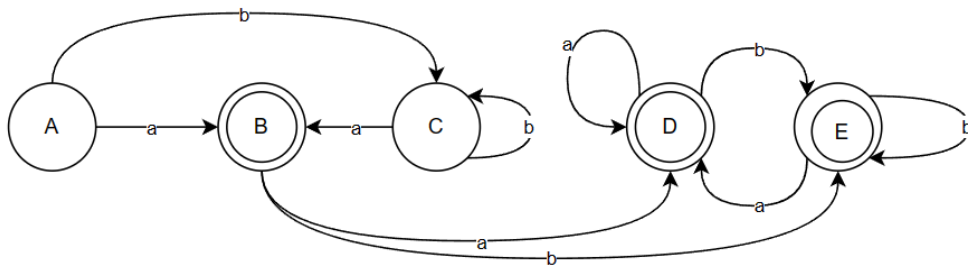
$$\begin{aligned} \text{move}(\{0, 1, 2, 3, 5, 6, 7, 8\}, b) &= \{5\} \\ \epsilon\text{-closure}\{5\} &= \{0, 1, 2, 3, 5, 6, 7, 8\} = C \\ \text{end} \end{aligned}$$

$$\begin{aligned} \text{move}(\{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17\}, a) &= \{4, 9, 14\} \\ \epsilon\text{-closure}\{4, 9, 14\} &= \{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17\} = D \\ \text{end} \end{aligned}$$

$$\begin{aligned} \text{move}(\{0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17\}, b) &= \{5, 15\} \\ \epsilon\text{-closure}\{5, 15\} &= \{0, 1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17\} = E \\ \text{end} \end{aligned}$$

## 1.3 使用你的 DFA 转移表将你的 NFA 转换为 DFA

根据上面创建的转换表，将 NFA 转换为 DFA，包含17的即为终止状态，图如下所示：



## 2. SLR

给定以下文法:

- (1)  $E \rightarrow X = Y$
- (2)  $X \rightarrow Xa$
- (3)  $X \rightarrow b$
- (4)  $Y \rightarrow cY$
- (5)  $Y \rightarrow d$

2.1 写出文法  $G$  的增广文法  $G'$ , 并根据该增广语法  $G'$  构造 LR(0) 解析的有穷自动机。

1. 增广文法  $G'$ :

$$E' \rightarrow E$$

$$E \rightarrow X = Y$$

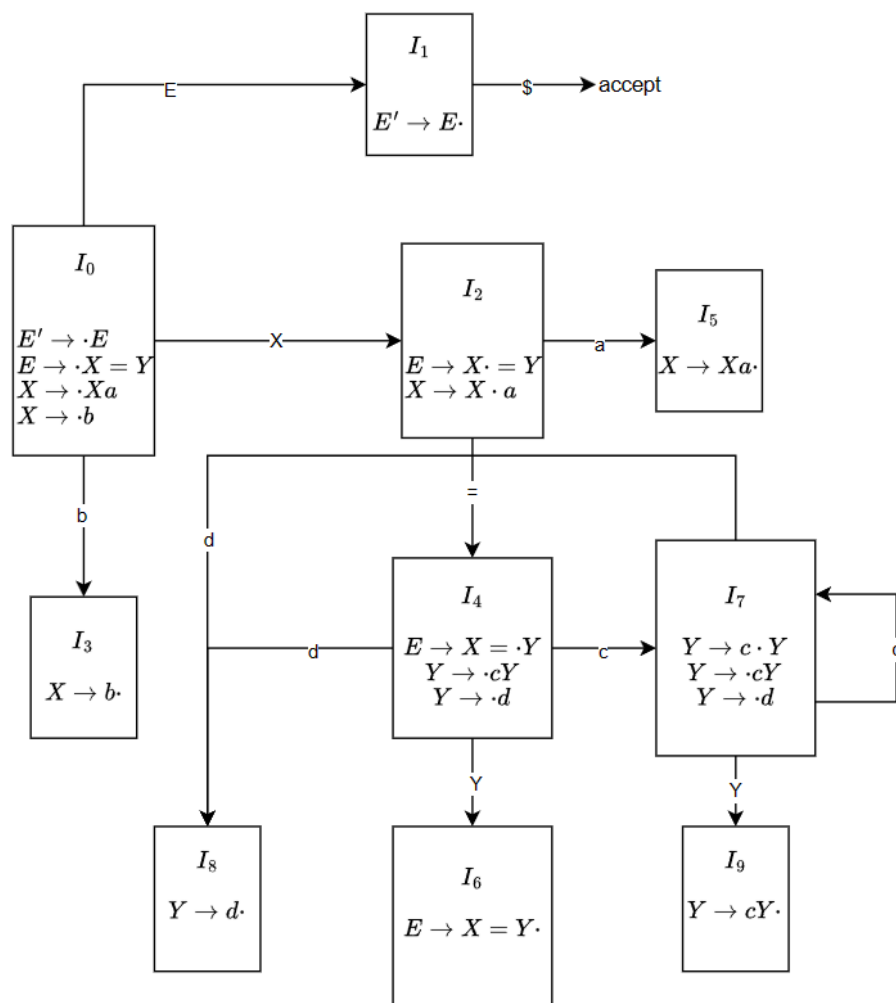
$$X \rightarrow Xa$$

$$X \rightarrow b$$

$$Y \rightarrow cY$$

$$Y \rightarrow d$$

## 2. LR(0) 解析的有穷自动机:



## 2.2 根据你画出来的有穷自动机构造 LR(0) 解析表

求解 LR(0) 解析表的过程如下:

### 1. Follow 集

- $\text{Follow}(E) = \{\$, \}$
- $\text{Follow}(X) = \{=, a\}$
- $\text{Follow}(Y) = \{\$, \}$

## 2. LR(0) 解析表

STATE	ACTION						GOTO		
	a	b	c	d	=	\$	E	X	Y
0		S3					1	2	
1						acc			
2	S5				S4				
3	r3				r3				
4			S7	S8					6
5	r2				r2				
6						r1			
7			S7	S8					9
8						r5			
9						r4			

---



## 2.3 列出使用 G 的 LR(0) 解析表解析输入串 baa = cd 的过程

STACK	SYMBOL	INPUT	ACTION
0		baa=cd\$	shift to state 4
03	b	aa=cd\$	reduce by (3) $X \rightarrow b$
02	X	aa=cd\$	shift to state 5
025	Xa	a=cd\$	reduce by (2) $X \rightarrow Xa$
02	X	a=cd\$	shift to state 5
025	Xa	=cd\$	reduce by (2) $X \rightarrow Xa$
02	X	=cd\$	shift to state 4
024	X=	cd\$	shift to state 7
0247	X=c	d\$	shift to state 8
02478	X=cd	\$	reduce by (5) $Y \rightarrow d$
02479	X=cY	\$	reduce by (4) $Y \rightarrow cY$
0246	X=Y	\$	reduce by (1) $E \rightarrow X = Y$
01	E	\$	accept