

## 编译原理 - 作业(1) : 词法分析

**Due:** Mar 23, 2023 at the beginning of the class (14:20)

1. 请描述以下正则表达式含义:

(1)  $((\epsilon|a)b^*)^*$

String of a's and b's.

(2)  $b(a|b)^*b$

String of a's and b's that start and end with b.

(3)  $(a|b)^*a(a|b)(a|b)$

String of a's and b's that the character third from the last is a.

(4)  $a^*ba^*ba^*ba^*$

String of a's and b's that only contains three b.

(5)  $(aa|bb)^*((ab|ba)(aa|bb)^*(ab|ba)(aa|bb)^*)^*$

String of a's and b's that has an even number of a and b.

2. 请对以下描述的字符串特征写出相应的正则表达式:

(1) 所有标识符以大写字母开始, 跟着 0 或多个字母或数字, 并以数字结尾。

$[A-Z][a-zA-Z0-9]^*[0-9]$

(2) 所有由按照字典顺序升序排列的小写字母构成的字符串。

$a^*b^*c^*\dots z^*$

(3) 所有由字母 a 和 b 构成但不包含 abb 的字符串。

$b^*(a|ab)^*$

(4) 所有由字母 a 和 b 构成但不包含连续 a 的字符串 (包括空串)。

$(a?b)^*a?$

(5) 基于字母表{a, b, c}的且字母 a 不出现在第一个 b 之后的所有字符串。

在评阅过程中发现这个题目存在歧义, 有两种理解。

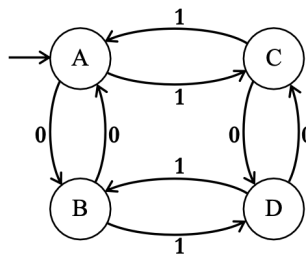
其中第一种理解是: 在出现第一个字母a之后, b就不允许再出现了。

其中第二种理解是: b只是不能紧跟这第一次出现的a出现

$(a|c)^*(b|c)^*$

$(a|c)^*b(c|b)(a|b|c)^*$

3. 设字母表{0, 1}上的有限自动机如下 (其中, A 是初始状态):



(1) 该自动机是否一个确定的有限自动机(DFA)? 为什么?

是一个 DFA, 因为它既没有空转移, 所有状态也不存在相同符号的多个射出弧。

- (2) 为让该自动机识别含有偶数个 1(包括零个 1)的所有串, 应将该自动机中的哪些状态改 为终结状态?

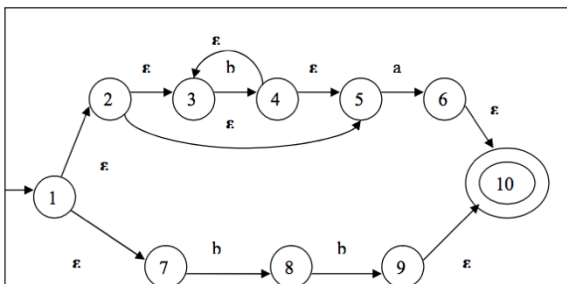
将状态 A 和状态 B 均改为终结状态。

- (3) 为让该自动机识别长度为奇数的所有串, 应将该自动机中的哪些状态改为终结状态?

将状态 B 和状态 C 均改为终结状态。

4. 考虑定义在字母表 $\Sigma = \{ a, b \}$ 正则语言:  $b^*a|bb$ :

- (1) 画出识别该语言的不确定有限自动机(NFA) ;



- (2) 将得到的 NFA 转化为等价的 DFA, 给出转换表和状态转换图 ;

转换过程 :

$\epsilon\text{-closure}(\{1\}) = \{1, 2, 3, 5, 7\}$

$\text{move}(\{1, 2, 3, 5, 7\}, a) = \{6\}$

$\epsilon\text{-closure}(\{6\}) = \{6, 10\}$

$\text{move}(\{1, 2, 3, 5, 7\}, b) = \{4, 8\}$

$\epsilon\text{-closure}(\{4, 8\}) = \{3, 4, 5, 8\}$

$\text{move}(\{6, 10\}, a) = \emptyset$

$\text{move}(\{6, 10\}, b) = \emptyset$

$\text{move}(\{3, 4, 5, 8\}, a) = \{6\}$

$\epsilon\text{-closure}(\{6\}) = \{6, 10\}$

$\text{move}(\{3, 4, 5, 8\}, b) = \{4, 9\}$

$\epsilon\text{-closure}(\{4, 9\}) = \{3, 4, 5, 9, 10\}$

$\text{move}(\{3, 4, 5, 9, 10\}, a) = \{6\}$

$\epsilon\text{-closure}(\{6\}) = \{6, 10\}$

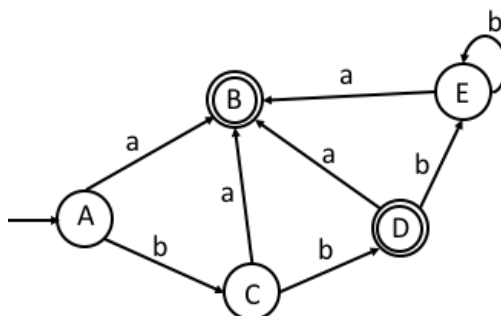
$\text{move}(\{3, 4, 5, 9, 10\}, b) = \{4\}$

$\epsilon\text{-closure}(\{4, 8\}) = \{3, 4, 5\}$

$\text{move}(\{3,4,5\}, a) = \{6\}$ 
 $\epsilon\text{-closure}(\{6\}) = \{6, 10\}$ 
 $\text{move}(\{3,4,5\}, b) = \{4\}$ 
 $\epsilon\text{-closure}(\{4,8\}) = \{3,4,5\}$ 

最终结果：

状态别名	DFA状态	a	b
A	$\{1,2,3,5,7\}$	$\{6,10\}$	$\{3,4,5,8\}$
B	$\{6,10\}$	-	-
C	$\{3,4,5,8\}$	$\{6,10\}$	$\{3,4,5,9,10\}$
D	$\{3,4,5,9,10\}$	$\{6,10\}$	$\{3,4,5\}$
E	$\{3,4,5\}$	$\{6,10\}$	$\{3,4,5\}$



- (3) 判断 (2) 中所得到的 DFA 状态是否已最小化。若是，请简述理由；若否，请将其最小化。

已经最小化。从 (2) 中得到的转换表可以看出，只有 D 和 E 在对输入 a/b 的转换是完全相同的，但这两个状态并不能合并，因为 D 是终结状态，而 E 不是。