

Chapter 2 Lexical Analysis

注意：

- 1、本次作业全部批改。
- 2、作业中全对以√标识；小问题在答案上做修改；大问题不作任何标记。
- 3、每位同学提交的本子都会有“阅”字和日期，并且作为平时成绩记录。
- 4、是按照多数同学共同存在的问题统计的，个别同学不具有代表性的问题没有统计。

Homework Analysis:

No. 2.1

a. $c^*a(a|c)^*b[a-c]^*$

问题的思考：定下第一个 a 和第一个 b 的位置，然后看看串 $(\sim a \sim b \sim)$ 中存在的可能及其表达法。

存在的问题：

- ✓ 很多同学考虑的太复杂，考虑到了 a, b 都不存在或者 a 存在 b 不存在的情况，写出了不存在 a 或者 b 的串，题目中说第一个 a 和第一个 b，那 a、b 必然存在。
- ✓ $(a|c)^*ac^*b[a-c]^*$ 也可以认为是正确的，其标识的 a 是 b 之前的最后一个 a
- ✓ 本题出错率：约 50%

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

问题的思考：将需要正则表达式描述的字符串进行如 $\dots(\sim a \sim a)(\sim a \sim a)(\sim a \sim a)(\sim)$ 的分割，便可以进行上述表达；或者进行如 $(\sim)(a \sim a \sim)(a \sim a \sim)(a \sim a \sim)\dots$ 的分割。

存在的问题：

- ✓ 很多人写成 $((b|c)^*a(b|c)^*a(b|c)^*)^*$ ，也正确。
- ✓ 本题出错率：约 50%

c. $(1|0)^*00$ 或 $(1(1|0)^*00)|0$

问题的思考：0、末两位为 0 的二进制数均是 4 的倍数。情形一：多位为 0 即为 0，0 打头的二进制数也可以；情形二：不能用多位 0 表示 0，不存在 0 打头的二进制数。

存在的问题：

- ✓ 问题不大。
- ✓ 出错率：很少，个别同学粗心。

d. $10101(0|1) \mid 1011(0|1)(0|1) \mid 11(0|1)(0|1)(0|1)(0|1) \mid (0|1)^*1(0|1)^*(0|1)(0|1)(0|1)(0|1)(0|1)$

问题的思考：大于 101001 的二进制数至少为六位二进制数，最小的大于 101001 的数为 101010，上面分四种情形来表示符合条件的二进制数。

存在的问题：

- ✓ 这个题目做出来的人很少，主要是分类方法没有考虑好。最后一项最后一项写成 $1(0|1)^*(0|1)(0|1)(0|1)(0|1)(0|1)(0|1)$ 也可以。

e. $(a|c)^*(b|bc(a|c)^*|ba|bac(a|c)^*)^*$ [仅供参考,如果发现问题,请同学们把自己认为

为正确的答案发到我的邮箱: tangdawei2009@gmail.com, 最好详细说明一下]

问题的思考: 先分析没有 b 的情况和一个 b 的情况, 一个 b 时, 可以单独的一个 b, 当后面紧跟一个 c 时, 可以为 $bc(a|c)^*$, 但后面跟一个 a 时, 可以为 ba 或者 $bac(a|c)^*$

存在的问题:

- ✓ 做出的太少, 比 d 题做出的还少, 主要是没有能够对问题进行清楚分类
- ✓ 有些同学做出来了, 但是忽略了开头的 $(a|c)^*$, 大部分同学不会。

f. $(00|0[1-7][0-7]^*)|(0[1-9][0-9]^*)$

问题的思考: 八进制数以 $0\times$ 的数, 八进制数的零为 00; 十进制数包含 0 以及不以 0 打头的十进制数。

存在的问题:

- ✓ 主要问题是很多人没有考虑到 0 的情况, 八进制的 0 为 00, 十进制的 0 为 0, 问题主要出在这里

g. $1|10$

问题的思考: $a^n + b^n = c^n$, 当 n 大于 2 时, 这个方程没有任何整数解。这就是数学史上著名的费马大定理或称费马最后的定理。

存在的问题:

- ✓ 做出来的不多, 可能主要是这个数学问题不理解, 知识面不够的缘故吧。

需要提醒的问题:

- ✓ 正则表达式写法不规范, 参见书中 P19~P20 的写法。有些我已经在作业上标注。
- ✓ 感觉有些同学的正则表达式写的没有错, 但是有表达的冗余, 关键在于对问题进行正确的分析与分类。
- ✓ d、e 两道题目多数没有写出来, 请大家注意分析问题和掌握思维方法, 分类的思想。
- ✓ 正则表达式怎么写的问题, 请参见每个题目的思考, 还可以给我来邮件。

No. 2.2

- a. 正则表达式的定义中可以看出, 没有任何运算是能够对字母进行计数的。
- b. 将回文串分为三段 $(1\sim n)(n+1)(n\sim 1)$, 从正则表达式的定义可以知道, 没有任何一种正则表达式运算能够将部分一的 $(1\sim n)$ 进行逆序排列并且表示出来。
- c. 存在括号、嵌套等文法, 无法由正则表达式来表达, 正则表达式不能表示。

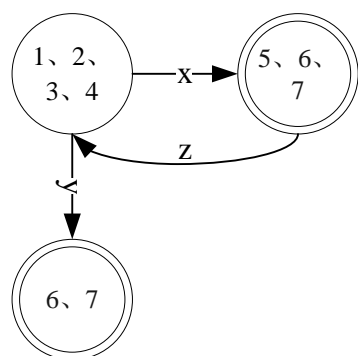
问题的思考: 本书中除了正则表达式的定义之外, 没有给出其它关于正则表达式的特征描述, 因此本题应该从定义出发来做说明。不用考虑的太复杂!

存在的问题:

- ✓ 多数同学直接采用“正则表达式不具备计数功能、记录功能”等结论来解释, 但本书中没有给出这样的结论, 因此不妥。
- ✓ 还有些同学通过 FA 与 RE 的关系来阐述正则表达式的表达能力, 两者间没有从分必要关系。
- ✓ 还有些同学用泵定理等进行证明, 超出本书范围, 不合适。

No. 2.5

- a.



本题问题不大，基本都能做对，也能够训练从 NFA 到 DFA 转化的过程。

b. 经过 31+步可以演算出这个 DFA，还是有些同学做出了结果：

$\text{edge}(1, \epsilon) = \{1\}$		
$\text{DFAedge}(\{1\}, a)$	$= \{1, 2, 3, 4, 5, 6\}$	←
$\text{DFAedge}(\{1\}, b)$	$= \{1\}$	←
$\text{DFAedge}(\{1, 2, 3, 4, 5, 6\}, a)$	$= \{1, 2, 3, 4, 5, 6\}$	←
$\text{DFAedge}(\{1, 2, 3, 4, 5, 6\}, b)$	$= \{1, 3, 4, 5, 6\}$	←
$\text{DFAedge}(\{1, 3, 4, 5, 6\}, a)$	$= \{1, 2, 4, 5, 6\}$	←
$\text{DFAedge}(\{1, 3, 4, 5, 6\}, b)$	$= \{1, 4, 5, 6\}$	←
$\text{DFAedge}(\{1, 2, 4, 5, 6\}, a)$	$= \{1, 2, 3, 4, 5, 6\}$	
$\text{DFAedge}(\{1, 2, 4, 5, 6\}, b)$	$= \{1, 3, 4, 5, 6\}$	
$\text{DFAedge}(\{1, 4, 5, 6\}, a)$	$= \{1, 2, 5, 6\}$	←
$\text{DFAedge}(\{1, 4, 5, 6\}, b)$	$= \{1, 5, 6\}$	←
$\text{DFAedge}(\{1, 2, 5, 6\}, a)$	$= \{1, 2, 6\}$	←
$\text{DFAedge}(\{1, 2, 5, 6\}, b)$	$= \{1, 3, 6\}$	←
$\text{DFAedge}(\{1, 5, 6\}, a)$	$= \{1, 2, 6\}$	
$\text{DFAedge}(\{1, 5, 6\}, b)$	$= \{1, 6\}$	←
$\text{DFAedge}(\{1, 2, 6\}, a)$	$= \{1, 2, 3\}$	←
$\text{DFAedge}(\{1, 2, 6\}, b)$	$= \{1, 3\}$	←
$\text{DFAedge}(\{1, 3, 6\}, a)$	$= \{1, 2, 4\}$	←
$\text{DFAedge}(\{1, 3, 6\}, b)$	$= \{1, 4\}$	←
$\text{DFAedge}(\{1, 6\}, a)$	$= \{1, 2\}$	←
$\text{DFAedge}(\{1, 6\}, b)$	$= \{1\}$	
$\text{DFAedge}(\{1, 2, 3\}, a)$	$= \{1, 2, 3, 4\}$	←
$\text{DFAedge}(\{1, 2, 3\}, b)$	$= \{1, 3, 4\}$	←
$\text{DFAedge}(\{1, 3\}, a)$	$= \{1, 2, 4\}$	
$\text{DFAedge}(\{1, 3\}, b)$	$= \{1, 4\}$	
$\text{DFAedge}(\{1, 2, 4\}, a)$	$= \{1, 2, 3, 5\}$	←
$\text{DFAedge}(\{1, 2, 4\}, b)$	$= \{1, 3, 5\}$	←
$\text{DFAedge}(\{1, 4\}, a)$	$= \{1, 2, 5\}$	←
$\text{DFAedge}(\{1, 4\}, b)$	$= \{1, 5\}$	←
$\text{DFAedge}(\{1, 2\}, a)$	$= \{1, 2, 3\}$	←
$\text{DFAedge}(\{1, 2\}, b)$	$= \{1, 3\}$	←
$\text{DFAedge}(\{1, 2, 3, 4\}, a)$	$= \{1, 2, 3, 4, 5\}$	←

DFAedge({1, 2, 3, 4}, b)	= {1, 3, 4, 5}	←
DFAedge({1, 3, 4}, a)	= {1, 2, 4, 5}	←
DFAedge({1, 3, 4}, b)	= {1, 4, 5}	←
DFAedge({1, 2, 3, 5}, a)	= {1, 2, 3, 4, 6}	←
DFAedge({1, 2, 3, 5}, b)	= {1, 3, 4, 6}	←
DFAedge({1, 3, 5}, a)	= {1, 2, 4, 6}	←
DFAedge({1, 3, 5}, b)	= {1, 4, 6}	←
DFAedge({1, 2, 5}, a)	= {1, 2, 3, 6}	←
DFAedge({1, 2, 5}, b)	= {1, 3, 6}	←
DFAedge({1, 5}, a)	= {1, 2, 6}	
DFAedge({1, 5}, b)	= {1, 6}	
DFAedge({1, 2, 3}, a)	= {1, 2, 3, 4}	←
DFAedge({1, 2, 3}, b)	= {1, 3, 4}	←
DFAedge({1, 3}, a)	= {1, 2, 4}	
DFAedge({1, 3}, b)	= {1, 4}	
DFAedge({1, 2, 3, 4, 5}, a)	= {1, 2, 3, 4, 5, 6}	
DFAedge({1, 2, 3, 4, 5}, b)	= {1, 3, 4, 5, 6}	
DFAedge({1, 3, 4, 5}, a)	= {1, 2, 4, 5, 6}	
DFAedge({1, 3, 4, 5}, b)	= {1, 4, 5, 6}	←
DFAedge({1, 2, 4, 5}, a)	= {1, 2, 5, 6}	
DFAedge({1, 2, 4, 5}, b)	= {1, 3, 5, 6}	←
DFAedge({1, 4, 5}, a)	= {1, 2, 5, 6}	
DFAedge({1, 4, 5}, b)	= {1, 5, 6}	
DFAedge({1, 2, 3, 4, 6}, a)	= {1, 2, 3, 4, 5, 6}	
DFAedge({1, 2, 3, 4, 6}, b)	= {1, 3, 4, 5}	
DFAedge({1, 3, 4, 6}, a)	= {1, 2, 4, 5}	
DFAedge({1, 3, 4, 6}, b)	= {1, 4, 5}	←
DFAedge({1, 2, 4, 6}, a)	= {1, 2, 3, 5}	
DFAedge({1, 2, 4, 6}, b)	= {1, 3, 5}	
DFAedge({1, 4, 6}, a)	= {1, 2, 5}	
DFAedge({1, 4, 6}, b)	= {1, 5}	
DFAedge({1, 2, 3, 6}, a)	= {1, 2, 3, 4}	
DFAedge({1, 2, 3, 6}, b)	= {1, 3, 4}	
DFAedge({1, 3, 6}, a)	= {1, 2, 4}	
DFAedge({1, 3, 6}, b)	= {1, 4}	
DFAedge({1, 2, 3, 4}, a)	= {1, 2, 3, 4, 5}	
DFAedge({1, 2, 3, 4}, b)	= {1, 3, 4, 5}	
DFAedge({1, 3, 4}, a)	= {1, 2, 4, 5}	
DFAedge({1, 3, 4}, b)	= {1, 4, 5}	
DFAedge({1, 4, 5, 6}, a)	= {1, 2, 5, 6}	
DFAedge({1, 4, 5, 6}, b)	= {1, 5, 6}	
DFAedge({1, 3, 5, 6}, a)	= {1, 2, 4, 6}	
DFAedge({1, 3, 5, 6}, b)	= {1, 4, 6}	
DFAedge({1, 4, 5}, a)	= {1, 2, 5, 6}	

DFAedge({1, 4, 5}, b) = ~~{1, 5, 6}~~

问题的思考：一种办法就是按部就班的将整个过程写出来，另外一种办法是实现该算法，让计算机去做这件事情。

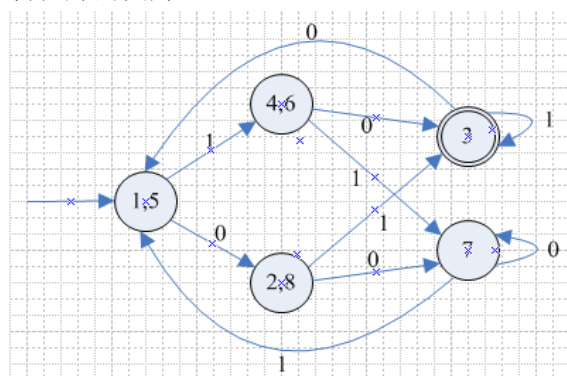
存在的问题：

- ✓ 多数同学感觉状态太多，没有能够给出完整的步骤和正确的结果。
- ✓ 没有对等价状态进行合并。
- ✓ 不过还是有个别同学做出来了

No. 2.6

等价的状态为 1 和 5，2 和 8，4 和 6

转化好的图为



本题问题不大，出错不多，出错情况主要在有些同学 1 和 5 的等价状态没有给出，或者 4 和 6 的等价状态没有给出。