

习题 3.1.1 写出表示下列语言的正则表达式：

- 字母表  $\{a, b, c\}$  上包含至少一个  $a$  和至少一个  $b$  的串的集合。
- 倒数第 10 个符号是 1 的 0 和 1 的串的集合。
- 至多只有一对连续 1 的 0 和 1 的串的集合。

$\exists 1-1(\infty)$      $\{a, b, c\}$  上串    以  $a$  和至少 1 个  $b$

$$\begin{aligned} & ab \neq ba \\ & \downarrow \quad \downarrow \\ \Rightarrow & (a+b+c)^* a (a+b+c)^* b (a+b+c)^* + (a+b+c)^* b (a+b+c)^* a (a+b+c)^* \\ (\Leftarrow) & (a+b+c)^* \left[ a (a+b+c)^* b + b (a+b+c)^* a \right] (a+b+c)^* \end{aligned}$$

$$(b+10+1)^* | (0+1)^*$$

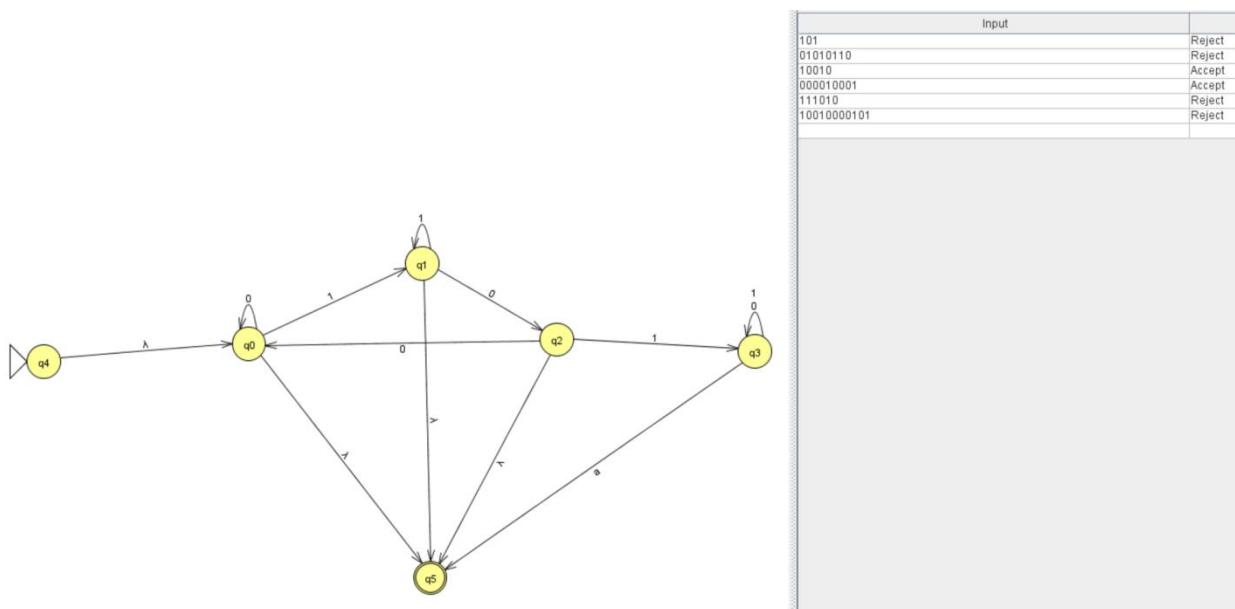
1 0 连续的 1  $\rightarrow$  有 1 1 或无连续的 1  
 0 有 - 对      没有 1 1 (2 对连续的 1)

$$\therefore (0+1)^* (11+1+\epsilon) (0+10)^*$$

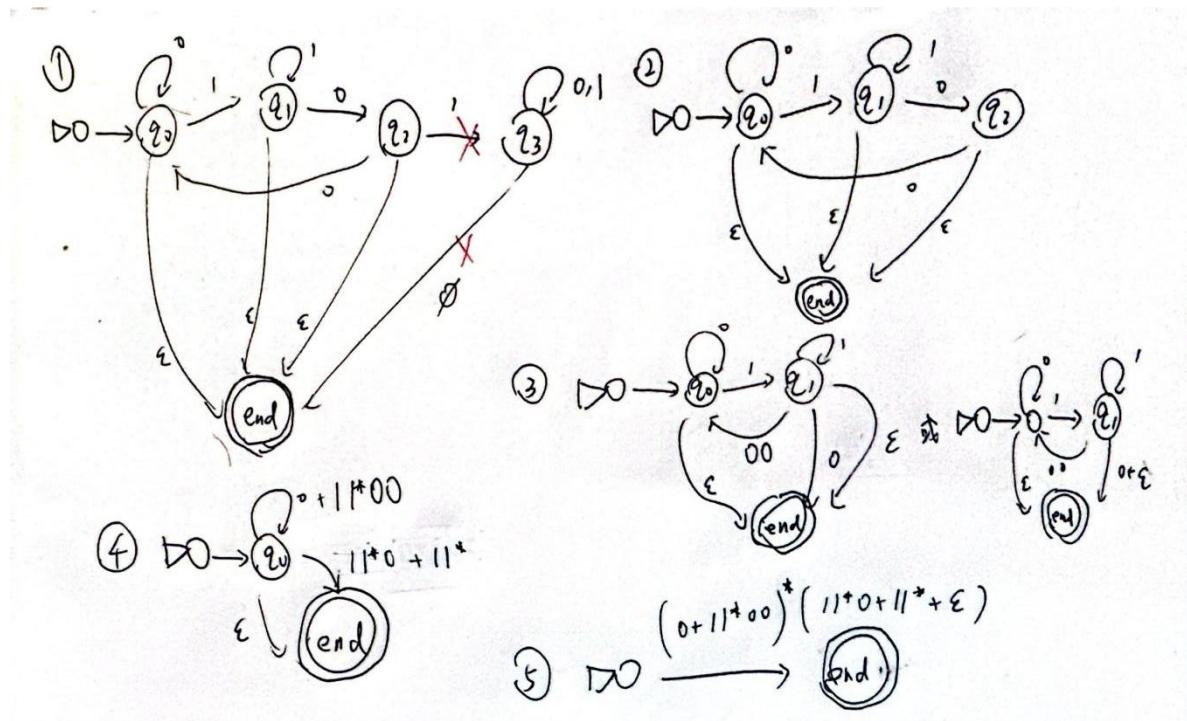
习题 3.1.3 写出表示下列语言的正则表达式：

- 不包含 101 作为子串的所有 0 和 1 的串的集合。

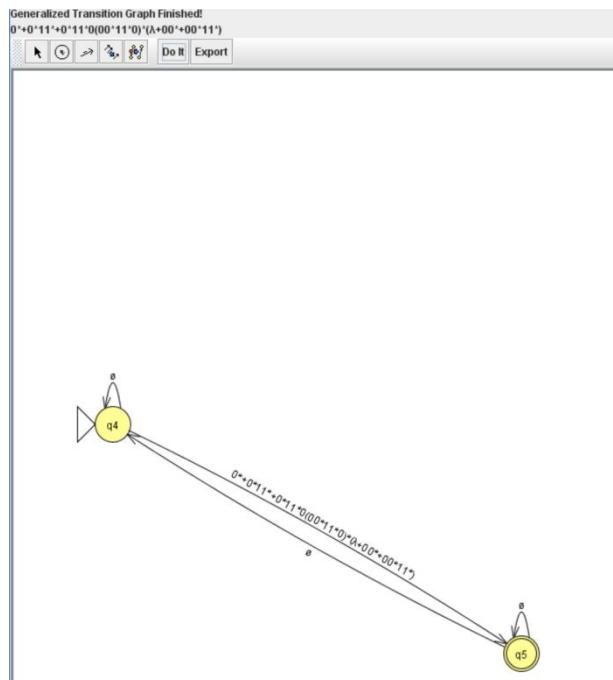
采用先画出 DFA 再进行转化的方法：画出对应的 DFA



其中  $q_3$  到  $q_6$  的是大非，也就是不可能到达最后的接受态。按照上课讲的第一种方法也就是 GNFA 方法就可以转化出来最终的正则表达式，转化的过程如下手写所示（省去  $q_3$  因为没有意义）：因此最终的等价的正则表达式为  $(0+11*00)*(11*0+11*+\epsilon)$



利用 JFLAP 可以给出他转化的正则表达式如下： $0^*+0^*11^*+0^*11^*0(00^*11^*)^*(\lambda+00^*+00^*11^*)$  与我们的表达式其实是等价的。（手写转化之前没有发现这个功能）

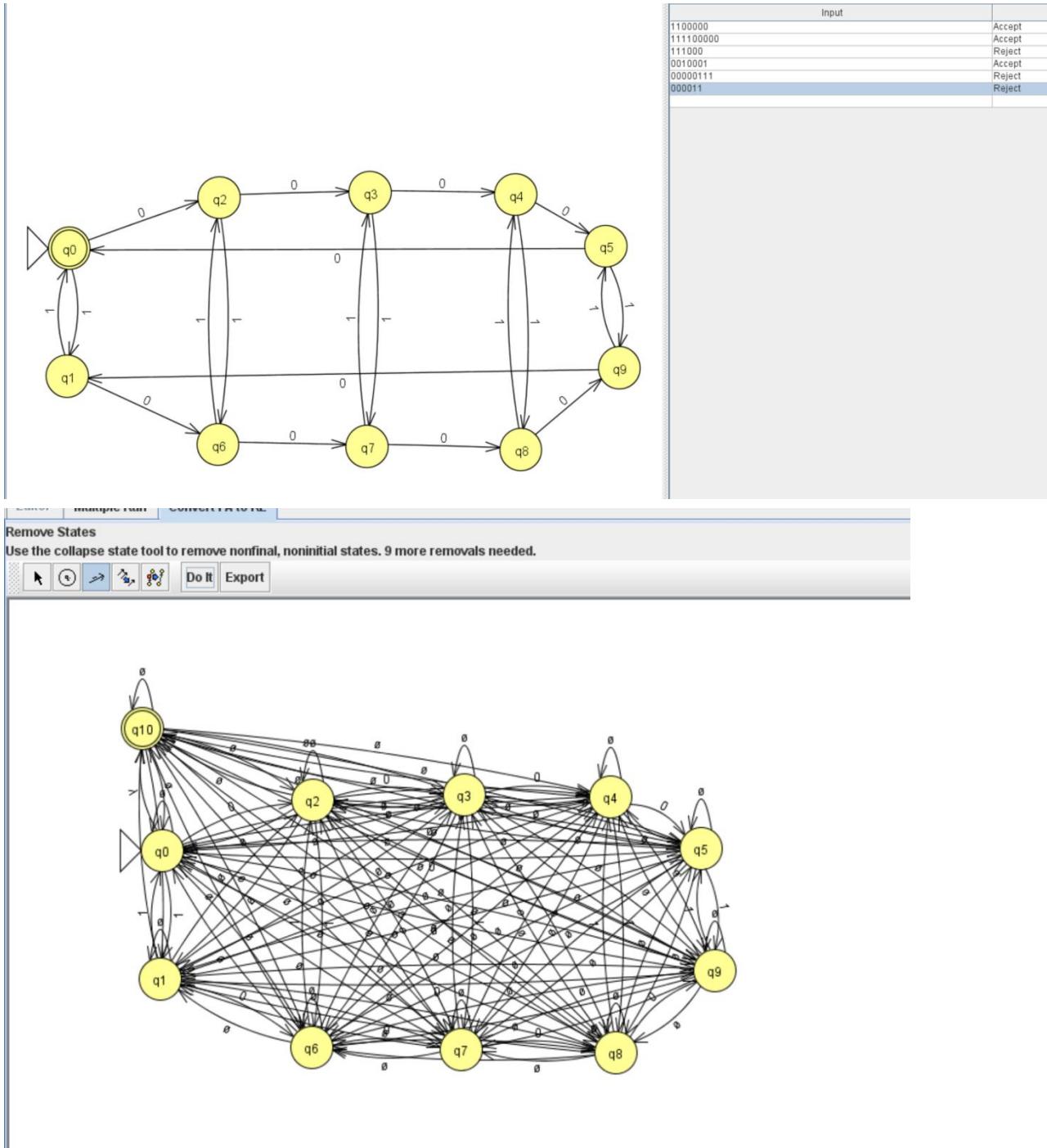


- b) 具有相同个数的 0 和 1，使得在任何前缀中，0 的个数不比 1 的个数多 2，1 的个数也不比 0 的个数多 2，所有这种 0 和 1 的串的集合。

因为 01 个数相同因此最先先到的就是 10 或者 01 组成的字符串，并且这俩正好满足任何前缀中 0 的个数不比 1 多 2, 1 的个数不比 0 多 2，举出例子，比如 1010、0101、1001、0110 等因此直接写就行：

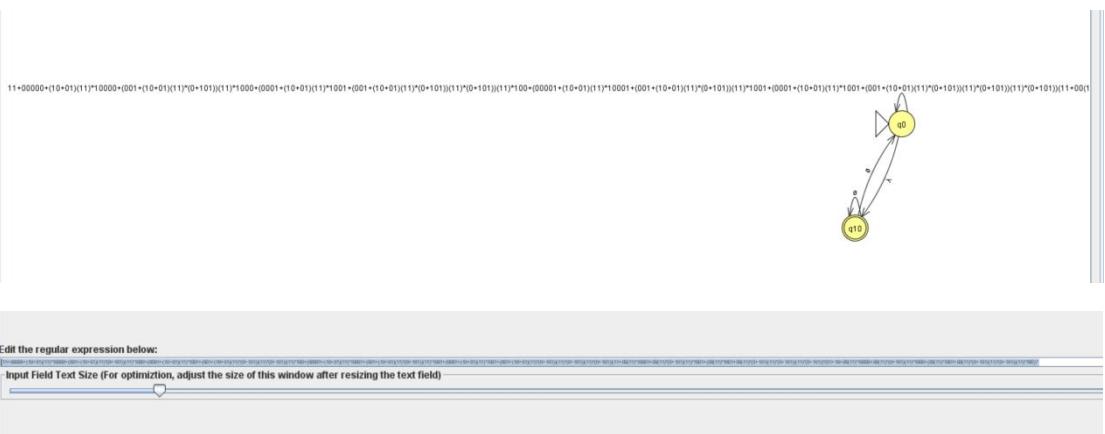
$(01+10)^*$

b) 0 的个数被 5 整除且 1 的个数是偶数的所有 0 和 1 的串的集合。



通过 JFLAP 自带的转化为 RE 的工具，转化出来如下的正则表达式：

```
(11+00000+(10+01) (11)*10000+(001+(10+01) (11)*(0+101)) (11)*1000+(0001+(10+01) (11)*1001+(001+(10+01) (11)*(0+101)) (11)*(0+101)) (11)*100+(00001+(10+01) (11)*10001+(001+(10+01) (11)*(0+101)) (11)*1001+(0001+(10+01) (11)*10001+(001+(10+01) (11)*(0+101)) (11)*(0+101)) (11)*1001+(0001+(10+01) (11)*1001+(001+(10+01) (11)*(0+101)) (11)*(0+101)) (11)*(0+101)) (11)*100+(00(11)*1001+00(11)*(0+101) (11)*100+(00(11)*1001+00(11)*(0+101) (11)*(0+101)) (11)*(0+101))*(01+10+00(11)*10000+00(11)*(0+101) (11)*1000+(00(11)*1001+00(11)*(0+101) (11)*(0+101)) (11)*(0+101)) (11)*100)*
```



**习题 3.1.4** 给出下列正则表达式语言的自然语言描述:

a)  $(1 + \varepsilon)(00*1)*0*$

不包含连续 1 的 01 字符串的集合

b)  $(0*1*)*000(0 + 1)*$

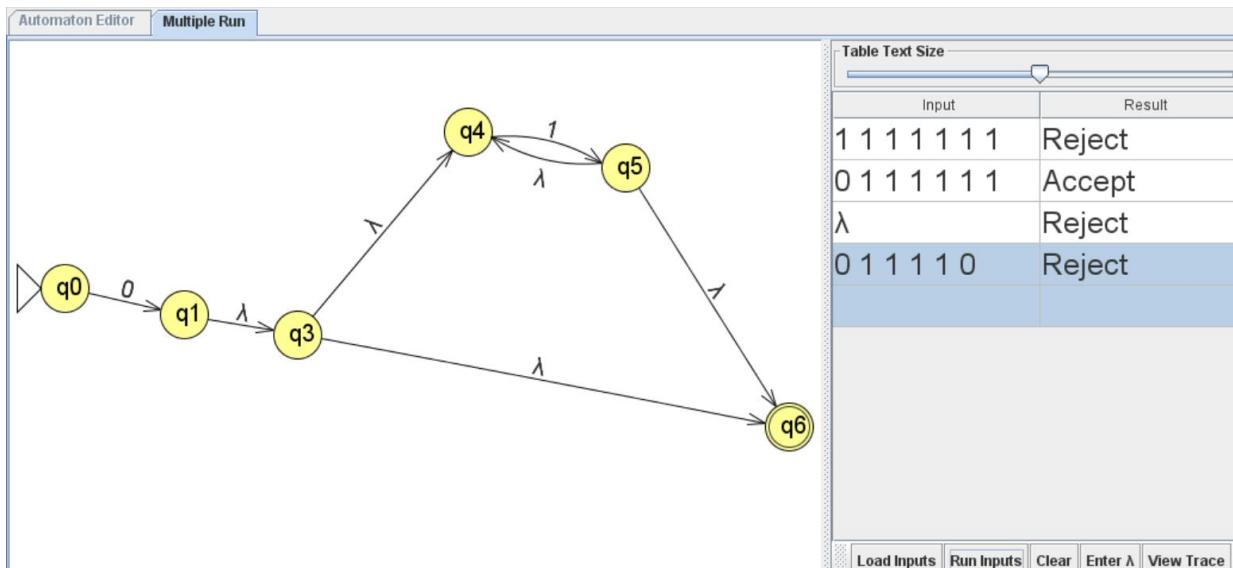
有 000 子串的 01 字符串的集合

c)  $(0 + 10)*1*$

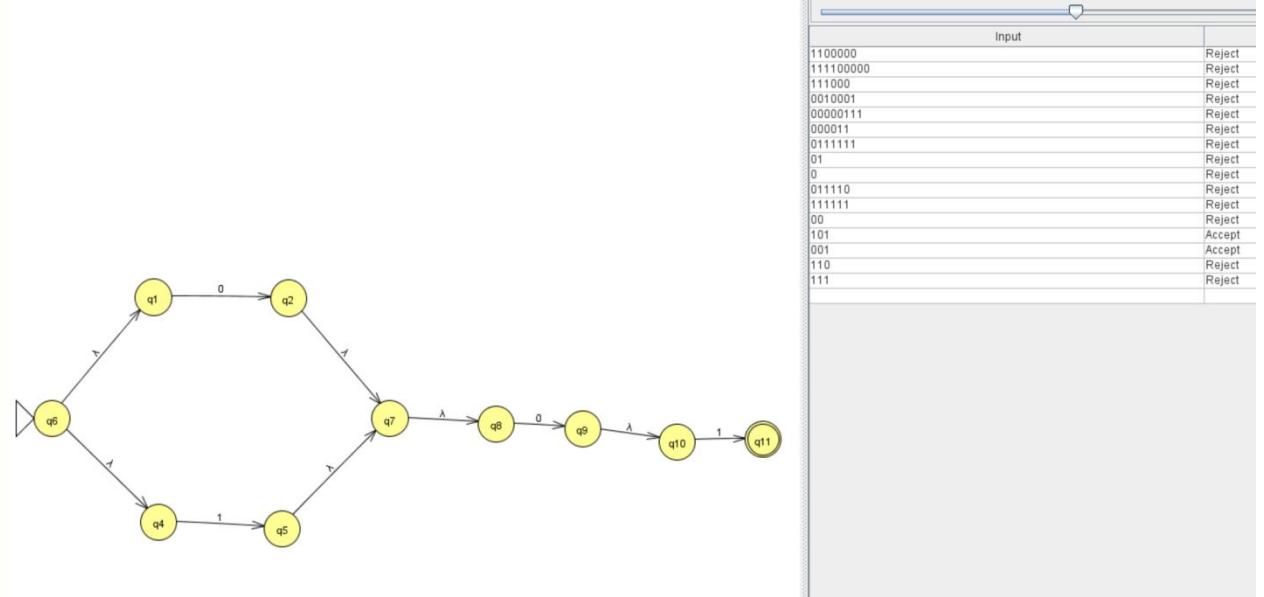
除了 01 字符串结尾以外不存在连续的 1 的字符串的集合

### 习题 3.2.4 把下列正则表达式转化成带 $\epsilon$ 转移的 NFA.

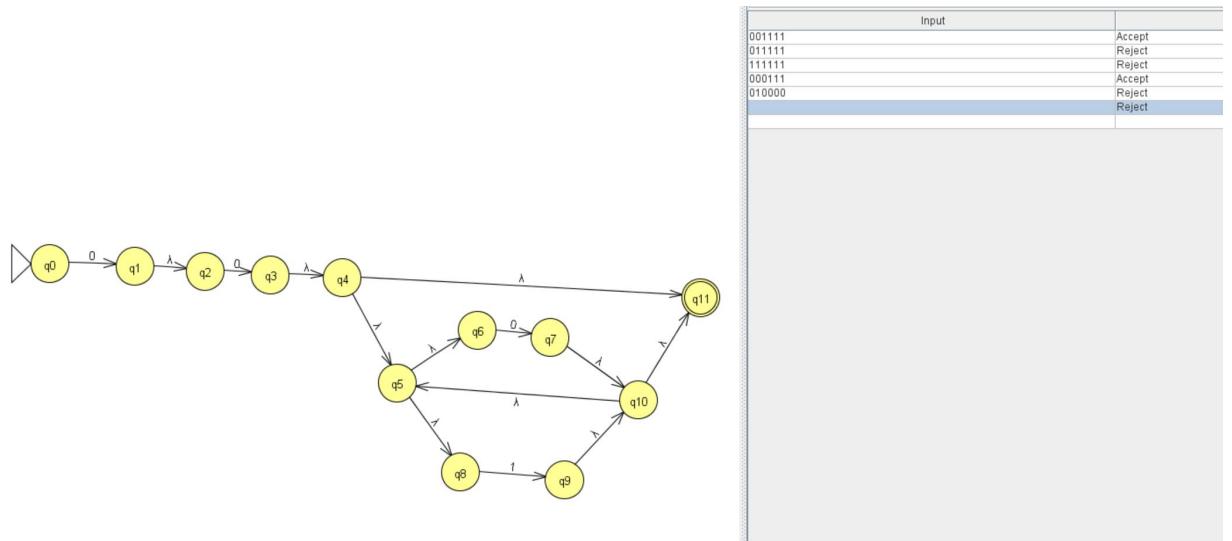
a)01\*



b) (0 + 1)01



c)  $00(0+1)^*$



习题 3.2.8 给出一个算法：输入一个 DFA  $A$ ，对于给定的  $n$ （与  $A$  的状态个数无关），

计算出  $AA$  所接受的长度为  $nn$  的串的个数。这个算法应当对于  $n$  和  $A$  的状态数来说都是多项式的。提示：使用定理 3.4 的构造所提示的技术。

3.28 设 DFA =  $A = \{q_0, q_1, q_2, q_3, q_4\}$ ,  $\Sigma = \{0, 1\}$

① 满足编码为  $l_1, l_2, \dots, l_s$  且  $0 \leq k \leq s$ ,  $i, j \in \Sigma$ ,  $l \geq 0$   
记  $N_{ij}^{(k)}(l)$  为从状态  $i$  到状态  $j$ , 经历  $k$  个字符  
长度为  $l$ , 且所有中间结点编码  $\leq k$  的路径数

①  $k=0$  不允许使用任何中间结点  
 ①)  $l=0$  且  $i=j$ ,  $N_{ii}^{(0)}=1$  记为空路径  $\emptyset$   
 ②)  $l=1$  且对于  $i, j$ ,  $N_{ij}^{(1)} = A_{ij} = \text{count}\left(\left\{a \in \Sigma : \delta(q_i, a) = j\right\}\right)$   
 (路径数)  $\circlearrowleft \circlearrowright$

$N_{ij}^{(k)}(l) = \begin{cases} 1, & i=j, l=0 \\ A_{ij}, & l=1 \\ 0, & i \neq j, l \geq 1 \end{cases}$

③  $k \geq 1$ , 则在所有由到到的路径中, 不包含编码号  $>k$  的结点  
故可分类讨论  
 ① 从未经过结点  $k$ , 记  $N_{ij}^{(k+1)}(l)$   
 ② 曾经过了  $k$  一次, 路径可分为以下的各路径  
 $i \xrightarrow{u_1} k \xrightarrow{u_2} k \xrightarrow{u_m} j \xrightarrow{u_{m+2}}$   
 $l_1, l_2, \dots, l_m$ , 且每个路径的中间结点  $\leq k-1$   
 $\sum_{l_i=0}^k l_i = l$   $\circlearrowleft \circlearrowright \circlearrowleft \circlearrowright$

每段选择相互独立, 则  
 $N_{ij} = \sum_{m \geq 2} N_{ik}^{(k+1)}(l_1) N_{kj}^{(k+1)}(l_2) \dots N_{kk}^{(k+1)}(l_{m-1}) N_{kj}^{(k+1)}(l_m)$   
 $\therefore N_{ij}^{(k)}(l) = N_{ij}^{(k+1)}(l) + \sum_{m \geq 2} \sum_{l_1=0}^k N_{ik}^{(k+1)}(l_1) \left( \prod_{t=2}^{m-1} N_{kt}^{(k+1)}(l_t) \right) N_{kj}^{(k+1)}(l_m)$

小节总结与表达式  $N_{ij}^{(k)}(n)$

共有 3 个概念, 因此对于一个字符串长度的时间复杂度为  $O(s^3)$  (字符串对每个路径度)  
 对每个状态对  $(i, j)$  需要对多  $s$  个中间状态和 1 个该过程的时间复杂度为  $O(s^3)$ , 需要对所有的  $n$  计算, 则遍历的总时间复杂度为  $O(n \cdot s^3)$

习题 3.3.2 给出一个正则表达式, 来表示在招聘广告中可能出现的薪水。考虑可能按小时、周、月或年发放的薪水。这些薪水可能有也可能没有\$ (如美元) 符号或其他单位 (如后面跟着的“K”)。可能有一个或多个邻近的单词标志着薪水。提示: 查看报纸上的分类广告或在线职位列表, 来获得些关于什么样的模式可能有用的想法。

招聘广告中的薪资表达方式十分多样, 通常包含货币符号 (如美元 \$、人民币 ¥、英镑 £、欧元 € 等), 并可能附带数值单位, 如千 (K)、百万 (M) 或十亿 (B) 等。除此之外, 薪资的支付周期也常有不同的表示形式, 例如按小时 (hour/hr)、天 (day)、周 (week/wk)、月 (month/mo/mth) 或年 (year/yr/annum/annual) 计算。

在实际招聘信息中, 薪资往往并非固定, 而是以区间形式呈现, 例如 “\$20–\$30/hr” 或 “\$50k–\$60k per year” 等。综合考虑货币符号、数值范围、单位缩写及多种表达习惯, 本题设计了一种如下风格的正则表达式, 用于识别招聘广告中的薪资信息:

$^\wedge (?i) (?:[\$€$

$\text{£¥}] | \text{USD} | \text{EUR} | \text{GBP} | \text{CNY} ) ? \backslash s * (? : (? : \backslash d \{1, 3\} (? : , \backslash d \{3\}) + | \backslash d+) (? : \backslash . \backslash d+) ? \backslash s * [\text{KMB}] ? ) \backslash s * (? : \backslash s * (? : \text{to} - | - ) \backslash s * (? : [\$€$

$\text{£¥}] | \text{USD} | \text{EUR} | \text{GBP} | \text{CNY} ) ? \backslash s * (? : (? : \backslash d \{1, 3\} (? : , \backslash d \{3\}) + | \backslash d+) (? : \backslash . \backslash d+) ? \backslash s * [\text{KMB}] ? ) ) ? \backslash s * (? : \backslash / | \backslash bp$   
 $\text{er} | \text{b} | \text{ba} (? : \text{n}) ? \text{b} ) ? \backslash s * (? : \text{hour} (? : \text{ly}) ? | \text{hr} | \text{h} | \text{day} | \text{d} | \text{week} | \text{wk} | \text{month} | \text{mo} | \text{mth} | \text{year} | \text{yr} | \text{annum} | \text{annua}$   
 $\text{l} | \text{daily} | \text{weekly} | \text{monthly} | \text{yearly} ) \$$

在 <https://regex101.com/> 测试结果如下：可见给出的样例均可以识别匹配：

The screenshot shows the regex101.com testing interface. At the top, a regular expression is defined:

```
:^(?i)(?:[$€£¥]|USD|EUR|GBP|CNY)?\s*(?:(?:(?:\d{1,3}(?:,\d{3})+|\d+)(?:\.\d+)?\s*[KMB]?)|\s*(?:to|-|)\s*(?:[$€£¥]|USD|EUR|GBP|CNY)?\s*(?:(?:(?:\d{1,3}(?:,\d{3})+|\d+)(?:\.\d+)?\s*[KMB]?)|\s*(?:\|\bper\b|\ba(?:n)?\b)?\s*(?:hour(?:ly)?\b|\hr|\h|day|\d|week|\wk|month|\mo|\mth|year|\yr|\annum|\annual|daily|\weekly|\monthly|\y|early)\$)
```

The "TEST STRING" section contains a list of sample strings, each followed by a green "Matched" indicator:

- \$20-\$30/hr
- CNY 12500 per month
- \$50,000 per year
- \$25/hour
- €12.5/hr
- 80k a year
- £400 per week
- \$20-\$30/hr
- \$1.2m annual
- EUR 30-35 per hour
- 2000 daily
- 45k monthly
- Salary: \$100,000
- 75k
- \$20-\$30
- \$40 per
- 12-to-18 students
- 20-30%
- +1-202-555-0100
- €,500 per year
- CNY 12,,000 per month
- 12 to 18 students
- 20-30%
- 11:00