

第 2 次作业-正则表达式-参考答案

4.1 分别写出下列正则表达式所代表的语言：

1) $(00+11)^*$

$\{\omega \mid \omega \in \{0, 1\}^*, \text{如果有则 } \omega \text{ 中 } 0, 1 \text{ 都是各自成对出现}\}$

2) $(0+1)^*0100^+$;

$\{\omega \mid \omega \in \{0, 1\}^*, \omega \text{以 } 010 \text{ 后接连续的 } 0 \text{ 结尾}\}。$

3) $(1+\epsilon)(00^*1)^*0^*$;

$\{\omega \mid \omega \in \{0, 1\}^*, \omega \text{不包含形如 } 11 \text{ 的子串}\}。$

4.2 写出下列语言的正则表达式：

1) $\{x \mid x \in \{0, 1\}^*, \text{且 } x \text{ 中不含形如 } 00 \text{ 的子串}\}。$

$(0+\epsilon)(11^*0)^*1^*$
 $1^*(01^+)^*(0+\epsilon)$
 $(1+01)^*(0+\epsilon)$
 $(0+\epsilon)(1+10)^*$

中间有 $(11^*0)^*$, $(1^+0)^*$, $(01^*1)^*$, $(01^+)^*$ 这样的串
 后再考虑首末字符为0, 1两种情况即可

2) $\{x \mid x \in \{0, 1\}^*, \text{且 } x \text{ 中含心形如 } 100 \text{ 的子串}\}。$

$(0+1)^*100(0+1)^*$

3) $\{x \mid x \in \{0, 1\}^*, x \text{ 包含偶数个 } 0 \text{ 和奇数个 } 1\}。$

$1+0(00+11)^*(01+10)$
 表示从状态偶数个0偶数个1
 到状态偶数个0奇数个1的转换

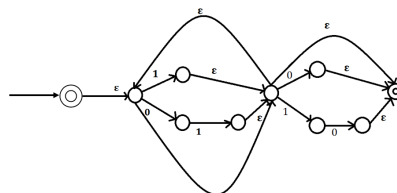
$(0(00+11)^*(01+10)+1)(00+11)^*((01+10)(00+11)^*(01+10)(00+11)^*)^*$

$(00+11)^*((01+10)(00+11)^*(01+10)(00+11)^*)^*$

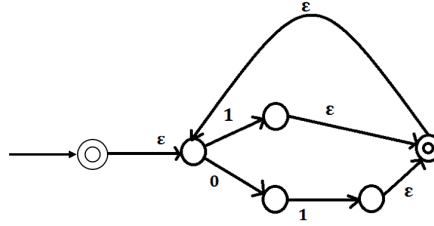
4.3 构造与下列正则表达式等价的有穷自动机：

1) $(1+01)^+ (\epsilon + 0+10) ;$

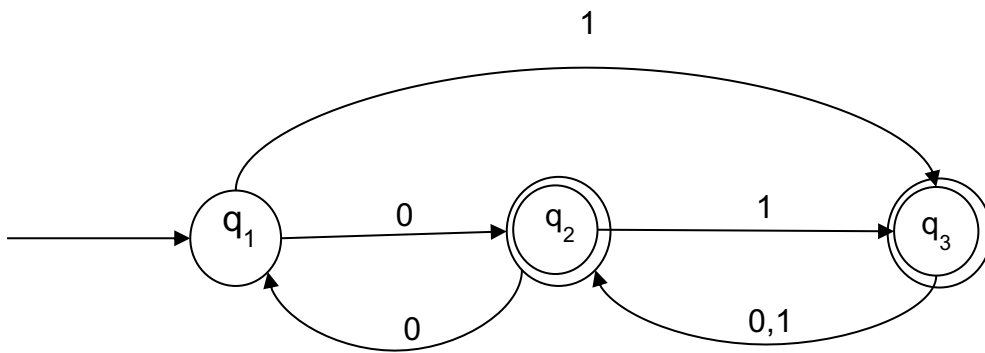
等价于
 $((00+11)+(01+10)(00+11)^*(01+10))^*$
 表示含有偶数个0偶数个1的自字符串



2) $(1+01)^+ ;$



4.4 给定下图所示的 DFA M，按照上课 PPT 中的证明方法二 (p34-40)，构造一个正则表达式表示 $L(M)$ 。



(此为课件例题)

1. 共 3 个状态， $n=3$; $k=0, 1, 2, 3$;
2. 根据递归公式求 r_{ij}^k

$$r_{ij}^k = r_{ik}^{k-1} (r_{kk}^{k-1})^* r_{kj}^{k-1} + r_{ij}^{k-1}$$

课件给出了部分:

	k=0	k=1	k=2
r_{11}^k	ϵ	ϵ	$(00)^*$
r_{12}^k	0	0	$0(00)^*$
r_{13}^k	1	1	0^*1
r_{21}^k	0	0	$0(00)^*$
r_{22}^k	ϵ	$\epsilon+00$	$(00)^*$
r_{23}^k	1	$1+01$	0^*1
r_{31}^k	\emptyset	\emptyset	$(0+1)(00)^*0$
r_{32}^k	$0+1$	$0+1$	$(0+1)(00)^*$
r_{33}^k	ϵ	ϵ	$\epsilon+(0+1)0^*1$

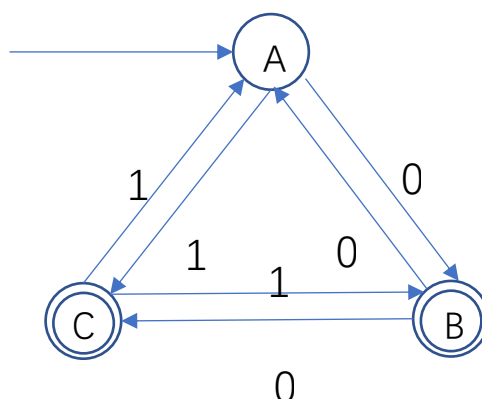
3. 由状态图可知 $L(M) = R_{12}^3 \cup R_{13}^3$ ，问题是求 $r_{12}^3 + r_{13}^3$ 代表 $L(M)$ 。

4. 故 $L(M) = r_{12}^3 + r_{13}^3 = r_{13}^2 (r_{33}^2)^* r_{32}^2 + r_{12}^2 + r_{13}^2 (r_{33}^2)^* r_{33}^2 + r_{13}^2$

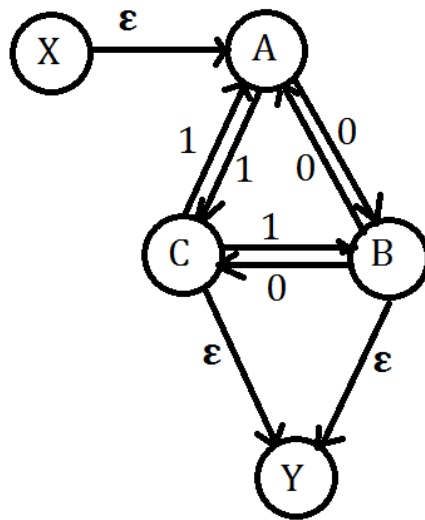
$$= 0^*1(\epsilon + (0+1)0^*1)^*((0+1)(00)^* + \epsilon + (0+1)0^*1) + 0^*1 + 0(00)^*$$

$$= 0^*1(\epsilon + (0+1)0^*1)^*((0+1)(00)^* + \epsilon) + 0(00)^*$$

4.6 请使用上课 PPT 中的两种证明方法（一题二解），构造下图所示的有穷自动机的正则表达式。

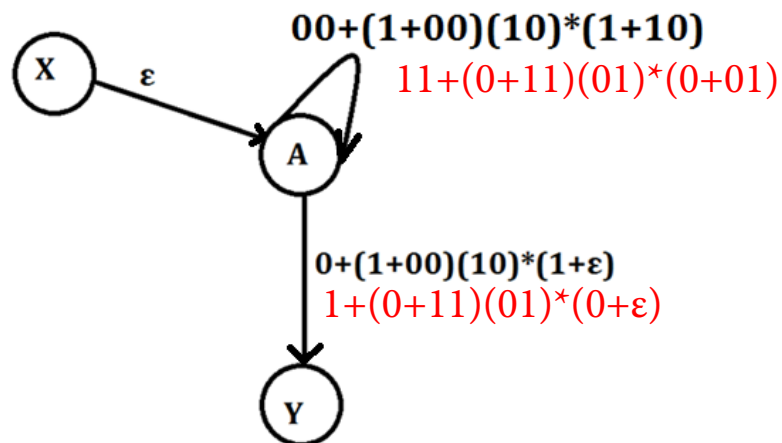


方法一: (1) 加首尾

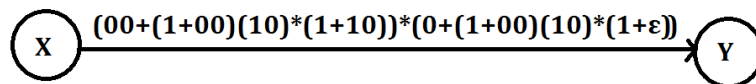


(2) 分别去掉状态 B、C

原答案顺序为依次去掉状态B、C；红色补充答案为依次去掉状态C、B



(3) 去掉状态 A



故正则表达式 $(00+(1+00)(10)^*(1+10))^*(0+(1+00)(10)^*(1+\epsilon))$ 就是所求。

方法二:

1. 共 3 个状态, $n=3$; 令 $A=1, B=2, C=3$, 故 $k=0, 1, 2, 3$;
2. 根据递归公式求 r_{ij}^k

$$r_{ij}^K = r_{ik}^{k-1} (r_{kk}^{k-1})^* r_{kj}^{k-1} + r_{ij}^{k-1}$$

3. 由状态图可知 $L(M) = R_{12}^3 \cup R_{13}^3$, 问题是求 $r_{12}^3 + r_{13}^3$ 代表 $L(M)$ 。
4. 故 $L(M) = r_{12}^3 + r_{13}^3 = r_{13}^2 (r_{33}^2)^* r_{32}^2 + r_{12}^2 + r_{13}^2 (r_{33}^2)^* r_{33}^2 + r_{13}^2 =$
 $(00+(1+00)(10)^*(1+10))^*(0+(1+00)(10)^*(1+\epsilon))$

4.7 给定若干文本文件（见附件），请使用正则表达式，检索且输出文件中的所有时间戳。要求：1、使用 Jupyter Notebook 编写；2、提交后缀为 ipynb 的源代码；3、检索结果用 print 输出。

```
In [4]: def search_dates(text):
patterns = [
    re.compile(r"(?:Jan|Feb|Mar|Apr|May|Jun|Jul|Aug|Sep|Oct|Nov|Dec)[a-z]*\.\?[s[0123]?d,?\s[12]\d{3}",
        re.IGNORECASE),
    re.compile(r"([01]?d/[12]\d{3})")
]
return list(itertools.chain.from_iterable([pattern.findall(text) for pattern in patterns]))
```

```
In [5]: dates = []
for file_name in os.listdir('data'):
    with open(os.path.join('data', file_name), 'r', encoding='utf-8-sig') as f:
        file = (' '.join([' '.join(line.strip().split()) for line in f]))
        dates.append(search_dates(file))
```

```
In [6]: print(dates[0])
```

#(?:Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec)	形如Jan Feb.....Dec这样的字符串
# [a-z]*	若干个英文字母
# \.	0个或1个 点.
# \s	1个空白字符
# [0123]?	0个或1个 0,1,2,3四个数字之一
# \d,?	0个或1个 一个数字后跟一个逗号
# \s	1个空白字符
# [12]	1个 1,2两个数字之一
# \d{3}	3个数字
# [01]?	0个或1个 0,1两个数字之一
# \d	1个数字
# /	1个斜杠
# [12]	1个 1,2两个数字之一
# \d{3}	3个数字

得分点:

两种匹配形式
 限定月份的英文拼写
 月份和日期的合法性检查