2152118 史君宝 形式语言与自动机 第二次作业

3.1.1
（A）至少包含一个 a 和一个 b，可知其他的是自由的。应该在 a 和 b 中穿插其他字母。
为

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

（B）倒数第 10 个字符是 1 的 0 和 1 的串的集合。可知前面的字符是随意的 0 或 1，第 10 个字符是 1，后面仍然是随意的 0 或 1，但是有个数限制为 9 个。
为

$$(0+1)^*1(0+1)^9$$

（C）至多只有一对连续的 1，所以我们有两种，先是有一对前面穿插，然后是没有一对的。
为

$$((1+\varepsilon)0^+)^*11(0^+(1+\varepsilon))^*+(1+\varepsilon)(0^+(1+\varepsilon))^*$$

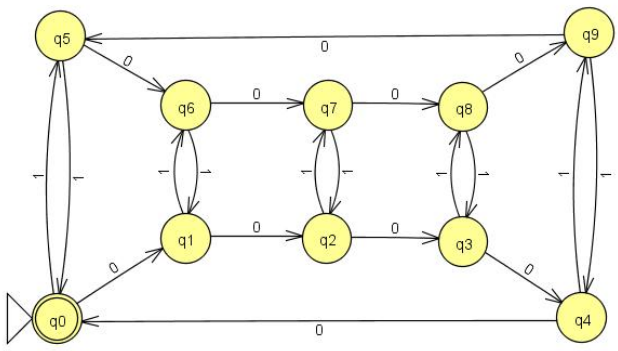
3.1.3
（A）不包含上述 101 集合的同时还要对 101 进行分解，使得不会出现 110 和 110 接连出现，使得前后连接成为 101
我们的思路是这样的，0 在串的中间出现的时候，只出现一次是极其危险的，但是在开头和结尾出现一次则没有什么问题。
为

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

（B）我们看到题目发现要求 0 和 1 的个数相同，这意味着两者同步出现才可能使得两者的个数相同，同时前缀中一个不可能比另一个多 2，也就是最小的子组应该两个成员，满足两者同步出现，且不会出现多 2
为

$$(01+10)^*$$

（C）这一题我没有比较好思路，因为可能的情况比较多，所以我决定从 FA 转为正则表达式，来求解。



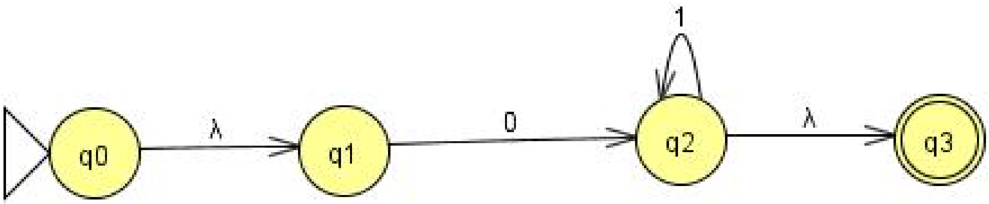
但是太麻烦了。我们将其分解为两种情况
$((1^a01^b01^c01^d01^e01^f)+(1^A01^B01^C01^D01^E01^F)(1^a01^b01^c01^d01^e01^f)^*(1^A01^B01^C01^D01^E01^F))^*+$
11
其中$(a+b+c+d+e+f) \bmod 2 = 0$;
$(A+B+C+D+E+F) \bmod 2 = 1$。

3.1.4
（A）识别所有不连续出现两个 1 的所有 0,1 串的语言。

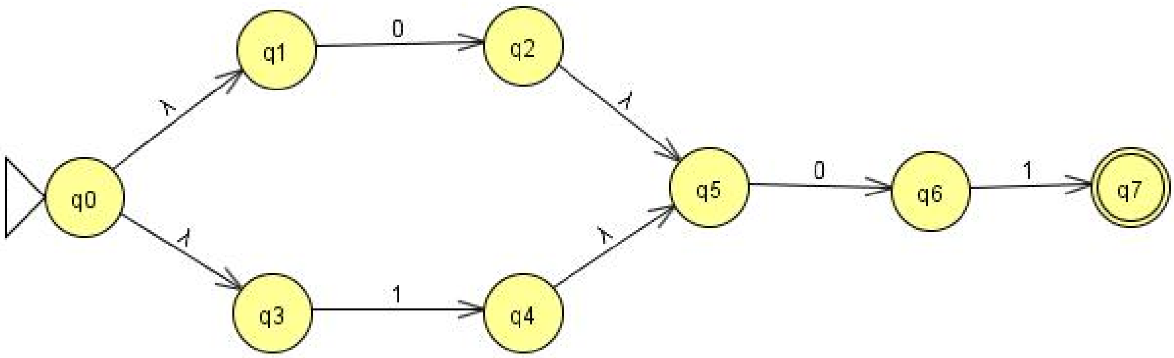
（B）识别串中有连续的 3 个 0 出现的所有 0,1 串的语言。

（C）识别串中有除了末尾出现连续的 1，其他地方不出现的 0,1 串的语言。

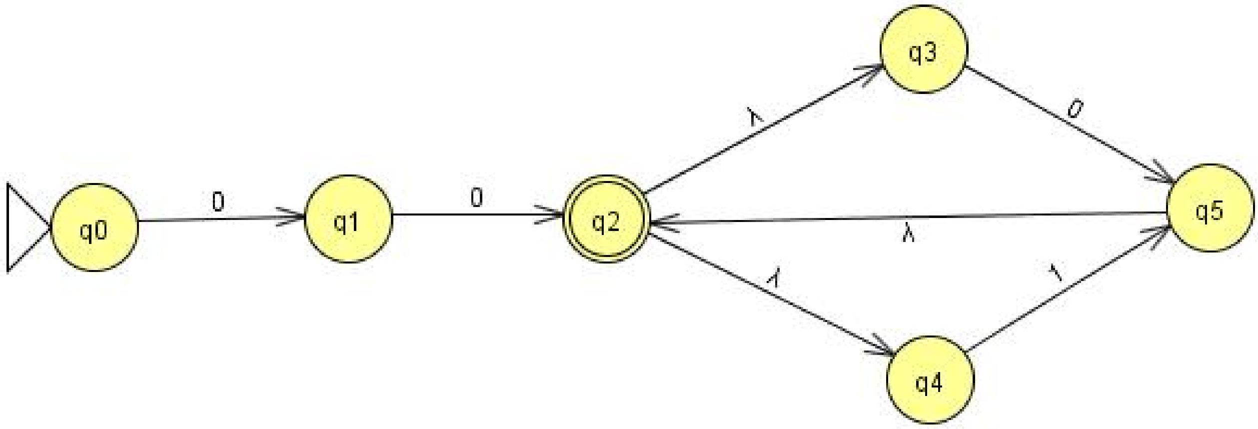
3.2.4
（A）



（B）



（C）



3.2.6

（A）原本上述的自动机所识别的语言是 L=L(A)，增加了从 q_f 到 q_0 的ε转移，那么识别的与语言就变成了

L^+

（B）原本上述的自动机所识别的语言是 L=L(A)，增加了从 q_0 到其所有可达的状态的ε转移，我们应该通过δ中找到 q_0 可达的状态，将这一部分修改一下就可以了

设 L 可以分解为 L=xw,即可以分解为 x 和 w 两个子串，其中 x 为单个字符。
上述状态中添加了 $\delta(q_0,\varepsilon)=\delta(q_0,x)$，也就是说是 L 去掉首字符的所有后缀子串。

（C)与上一题不同的是,这一题中我们添加的是所有能够到达 q_f 的状态的ε转移，原本上述的自动机所识别的语言是 L=L(A)。

设 L 可以分解为 L=xw,即可以分解为 x 和 w 两个子串。
上也就是说是 L 去掉尾部子串的所有前缀子串。与上一题不同的是，这里去掉的是尾部子串而非尾部字符，所以也可以识别空串。

（D）结合前两问的结果可知
上述可以识别 L 的所有子串，无前缀后缀要求，可以为空串。

3.2.7

上面的问题，可知三个变化都可以单独使用，但是变化（3）并不能和其他一起使用，也就是说，可以使用的子集是{(1)}{(2)}{(3)}{(1)(2)}，当两个自动机的闭包共用状态时，可能会发生错误。

3.2.8

我们仿照课本上的定理 3.4 来求解这一题。
用 $R^k ij(n)$ 来表示正则表达式,与课本上的定义是相同,并设 k 为所有状态数的数

量总和。
当经过的状态数不大于 K 的时候，最终的状态必定递归到 0，
$R^0ij(0)=1$,我们将对应的所有 $R^0ij(n)$也对应为 $R^0ij(0)=1$
递归公式有：$R^kij(n)=R^{k-1}ij(n)+R^{k-1}ik(x1)R^{k-1}kk(x2)\cdots\cdots R^{k-1}kk(xm-1)R^{k-1}kj(xm)$，上述的公式表示从 i 到 j 的过程中经过 x1 到 xm 个 K 状态，上述递归就可以得到结果。
同样，上述递归的数量是从 K 递归到 0，同时过程中经过的 K 状态也是与串的长度 n 有关，不超过 n，所以说上述公式的复杂度与 k 和 n 是多项式复杂度的。

3.3.1
（‘+’+ε）在国内为空，国外为‘+’
之后是国际冠字，一般根据国家分为 1-3 位，所以为
$(0+1+2+3+4+5+6+7+8+9)^1+(0+1+2+3+4+5+6+7+8+9)^2+(0+1+2+3+4+5+6+7+8+9)^3$
然后是国家内的区号，一般为 1-4 位
$(0+1+2+3+4+5+6+7+8+9)^1+(0+1+2+3+4+5+6+7+8+9)^2+(0+1+2+3+4+5+6+7+8+9)^3+(0+1+2+3+4+5+6+7+8+9)^4$
后面是本地号码，一般是 8-11 位
$(0+1+2+3+4+5+6+7+8+9)^8+(0+1+2+3+4+5+6+7+8+9)^9+(0+1+2+3+4+5+6+7+8+9)^{10}+(0+1+2+3+4+5+6+7+8+9)^{11}$

所以格式为
（‘+’+ε）$((0+1+2+3+4+5+6+7+8+9)^1+(0+1+2+3+4+5+6+7+8+9)^2+(0+1+2+3+4+5+6+7+8+9)^3)((0+1+2+3+4+5+6+7+8+9)^1+(0+1+2+3+4+5+6+7+8+9)^2+(0+1+2+3+4+5+6+7+8+9)^3+(0+1+2+3+4+5+6+7+8+9)^4)((0+1+2+3+4+5+6+7+8+9)^8+(0+1+2+3+4+5+6+7+8+9)^9+(0+1+2+3+4+5+6+7+8+9)^{10}+(0+1+2+3+4+5+6+7+8+9)^{11})$

3.3.2
前面的内容（a+b+c+……+z+“汉字”）
钱的符号在前（$+¥+K+€+ε)
具体金额 $(0+1+2+3+4+5+6+7+8+9)^*$
钱的符号在后（$+¥+K+€+ε)
关于时间（“年”+“月”+“日”+“时”+“year”+“month”+“day”+“hour”）
后面的内容（a+b+c+……+z+“汉字”）

(a+b+c+……+z+“汉字”)($+¥+K+€+ε)$(0+1+2+3+4+5+6+7+8+9)^*$($+¥+K+€+ε)(“年”+“月”+“日”+“时”+“year”+“month”+“day”+“hour”)(a+b+c+……+z+“汉字”)