习题 5.1.5 设 $T = \{0, 1, (,), +, *, \phi, e\}$，可以把 T 看作字母表为 $\{0,1\}$ 的正则表达式所使用的符号的集合，惟一的不同是用 e 来表示符号 ε，目的是为了避免有可能出现的混淆。你的任务是以 T 为终结符号集合来设计一个 CFG，该 CFG 生成的语言恰好是字母表为 $\{0,1\}$ 的正则表达式。

5.1.5　$G = (V, T, P, S)$ ← 开始字符

变元集　终结符集　产生式规则

$\{0,1\}$ 的正则表达式

$T = \{0, 1, (,), +, *, \phi, e\}$

$V: S$

$P =$
$S \to S+S$　$S \to (S)$
$S \to S^*$　$S \to SS$
$S \to 0$　$\therefore S \to S+S \mid S^* \mid SS \mid (S) \mid 0 \mid 1 \mid \phi \mid e$
$S \to 1$
$S \to \phi$
$S \to e$

习题 5.4.7 下面的文法生成的是具有 x 和 y 操作数、二元运算符 $+$、$-$ 和 $*$ 的前缀表达式：

$$E \to +EE \mid *EE \mid -EE \mid x \mid y$$

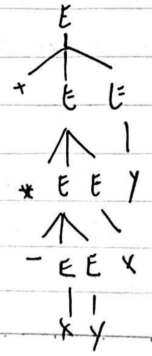
a) 找到串 `+*-xyxy` 的最左推导、最右推导和一棵语法分析树。

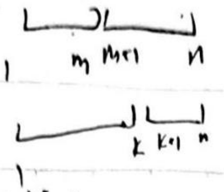
b) 证明这个文法是无歧义的。

5.4.7　$E \to +EE \mid *EE \mid -EE \mid x \mid y$

a) 最左推导　$E \to +EE \to +*EEE \to +*-EEEE \to +*-xEEE$
$+*-xyxy$　$\to +*-xyEE \to +*-xyxE \to +*-xyxy$

最右推导　$E \to +EE \to +Ey \to +*EEy \to +*Exy \to +*-EExy$
$\to +*-Eyxy \to +*-xyxy$





(b) 证明这个文法无歧义

无歧义：语法树唯一 → + - *　　记为命题1

① 证明 $\forall w \in L(G)$，w中符号的数量比x,y的数量之和少1，且w中任何后缀中，+ * - 的数量严格少于x,y的数量之和，记为命题2

基础：$|w|=1$ 时，$E \to x$ 或 $E \to y$ ∴ $w=x$ 或 $w=y$　$|w|=1$ 符合命题1,2

归纳：$|w|>1$ 时，推理的最后一步一定使用了 $E\to +EE$, $E\to -EE$, $E\to *EE$，不妨设为 $E \to +EE$ 则 $w=+w_1w_2$　w_1,w_2 满足 $E \overset{*}{\Rightarrow} w_1$　$E \overset{*}{\Rightarrow} w_2$

则对于 $w_1\ w_2$ 通过归纳法也可通过不断的递归推理可知 $w_1\ w_2$ 满足命题1,2

$w_1 \to +w_{11}w_{12}$

$w_{11} \to +w_{111}w_{112}$　$w_{12} \to +w_{121}w_{122}$

...

$|w|=1$ 时 $w=x,y$

则 $w=+w_1w_2$，其中w中 +,-,* 的数量比x,y少1

同理易得，由于 w_1w_2 的后缀中 +,*,- 数量严格少于x,y数量

则 w中，w的后缀中 +,*,- 数量小于x,y数量

② 证明 $w=+w_1w_2$ 的划分唯一

$|w|=1$ 时，$E\Rightarrow x$ 或 $E \Rightarrow y$，有唯一语法树

$|w|>1$ 时，则推导的第1步一定使用了3个式 $E\to +EE$，$E\to -EE$，$E \to *EE$ 之一 不妨设为 $E\to +EE$，$w=+w_1w_2$

若w有2种不同的划分方法 F_1，F_2

F_1 中 $w_1=w_{[1,m]}$　$w_2=w_{[m+1,n]}$　F_2 中 $w_1=w_{[1,k]}$　$w_2=w_{[k+1,n]}$

设 $k>m+1$，由命题1，w中 +,-,* 的数量比x,y数量和少1

则 $w_{[m+1,k]}$ 中 +,*,- 的数量必须与x,y的数量相等

这与命题2矛盾，故 $w=+w_1w_2$ 的划分唯一

综上①②，可知该文法具有无歧义性

习题 5.2.2 假设 G 是一个 CFG，并且它的任何一个产生式的右边都不是 ε。如果 w 在 $L(G)$ 中，w 的长度是 n，w 有一个 m 步完成的推导，证明 w 有一个包含 $n+m$ 个节点的分析树。

设CFG，且它的任何一个产生式的右边都不是 ε，若 $w\in L(G)$ $|w|=n$，

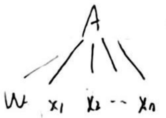
w有一个m步完成的推导，证：w有一个包含 $n+m$ 个节点的分析树

思路：1次推导 → 增加一个内部节点

1个终结符 → 叶节点

，推导使用k步的所有字符串，其分析树节点数 = 字符串长度 + k

① $k=1$时，$A\Rightarrow w$，说明此时存在产生式 $A\to w$，$|w|=n$，w由n个终结符构成，此时节点 $=1+n=n+k$



$x_1 \dots x_n$ 为终结符

x_i可能相同或不相同

∴只作次序说明

② 归纳，假设对∀ k步推导得出的子符串w，都有$|w|+k$个节点

要证：若用了$k+1$步推导得到了w，那么对应的分析树有$|w|+m+1$个节点

$|w|=n$

设最后一步使用了产生式 $A\Rightarrow X_1X_2\dots X_m$

A是根，最后一步 抽象为 将A展开为多个符号 $X_1\dots X_m$，此X_i都能生成 1个子串 w_i

$w=w_1w_2\dots w_k$，设$|w_i|=|n_i|$，t_i为X_i生成w_i所需的推导步数

注意推导步数$k+1$，根的展开1步，剩下k步给子树

$\Sigma t_i=k$

对∀ X_i，若X_i为终结符x_i，则 $w_i=x_i$ $|w_i|=|n_i|=1$，$t_i=0$

若X_i是非终结符，则X_i生成w_i要t_i步推导，由假设可知，

该子树的节点总数 $=n_i+t_i$

$A\to X_1\dots X_m$

$\Sigma t_i = k+1-1 = k$

∴总节点数 $=$ 1+子树节点之和 $=1+\Sigma(n_i+t_i)$

$=1+\Sigma n_i+\Sigma t_i$

$=1+|w|+k$ ∴得证

习题 5.2.3 假设在习题 5.2.2 中除了 G 中可能有右端为 ε 的产生式外其他所有的条件都满足，证明此时 w（w 不是 ε）的语法分析树有可能包含 $n+2m-1$ 个节点，但不可能更多。

G中允许有右部为 ε 的产生式，树有可能包含 $n+2m-1$ 节点，但不可能更多

证明：若某变量 A，经过 k 步推导之后，得到串 u，则 $\exists$ 分析树，其节点数最多为 $|u|+2k-1$，C 为分析树节点数

① $k=1$ $A \Rightarrow u$ 用1步得到 $u=x_1 \cdots x_n$，x_i 为终结符 $n=|u|$

节点数 $= 1+n = n+1$

$|u|+2k-1 = n+2\times1-1 = n+1$，成立

$A \Rightarrow u$ 且 $u=\varepsilon$ 时，$|u|=0$，节点数 $=1$ $|u|+2k-1 = 0+2\times1-1=1$，成立

② 假设，某变量用 r 步推导出串 v，就有分析树节点 $\leq |v|+2r-1$

设某变量 A，用 $k+1$ 步推导出串 W

推导最后一步为 $A \Rightarrow X_1X_2\cdots X_t$（$X_i$ 可能是终结符或不是）

$W = w_1w_2\cdots w_t$

① 若 X_i 是终结符 $w_i = X_i$，$n_i=|w_i|=1$，$r_i=0$

② 若 X_i 不是终结符，则从 X_i 推导出 w_i，$|w_i|=n_i$，步数为 r_i

$A \Rightarrow X_1\cdots X_t$ 占1步，则 $\sum_{i=1}^{t} r_i = k$ 且 $|W| = n = \sum_{i=1}^{t} n_i$

设分析树中以 X_i 为根的子树节点数记为 c_i

① X_i 为终结符 $r_i=0$ $w_i=X_i$ $n_i=1$ $c_i=1=n_i$

$\therefore c_i \leq n_i = n_i + 2r_i$

② X_i 为非终结符， $r_i \geq 1$ $X_i \xRightarrow{r_i} w_i$

由假设可知 $c_i \leq n_i+2r_i-1$

记 $\delta_i = \begin{cases} 1, & X_i\text{为非终结符} \\ 0, & X_i\text{为终结符} \end{cases}$ 由①② $c_i \leq n_i+2r_i-\delta_i$

设 $j = \sum_{i=1}^{t} \delta_i$，即为 X_i 中非终结符个数

$$C = 1+\sum_{i=1}^{t} c_i \leq 1+\sum_{i=1}^{t}(n_i+2r_i-\delta_i) = 1+\sum n_i+2\sum r_i-\sum\delta_i$$

$$= 1+n+2k-j \leq n+2m-1$$

$j=0$ 即 X_i 均为终结符，$\therefore$ 成立

1. 对于下列语言，分别构造接受它们的 PDA：

1） $\{0^n1^m \mid n \geqslant m \geqslant 1\}$

2） $\{1^n0^n1^m0^m \mid n, m \geqslant 1\}$

3） 含有 0 的个数和 1 的个数相同的所有 0，1 串

（1）这里对第一题进行详细的实验以及验证，包括三种方法，，剩下两题不进行详细的解释，以做出来为准。第一种方法如下：

(1) $\{0^n1^m \mid n \geqslant m \geqslant 1\}$

$0^n1^m = 0^{n-m}0^m1^m$

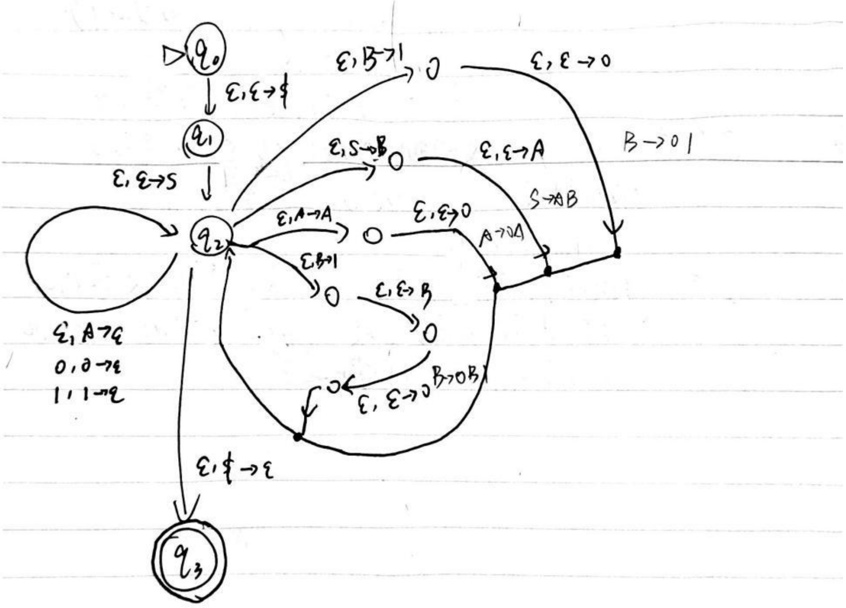
构造 CFG G = (V, T, P, S)

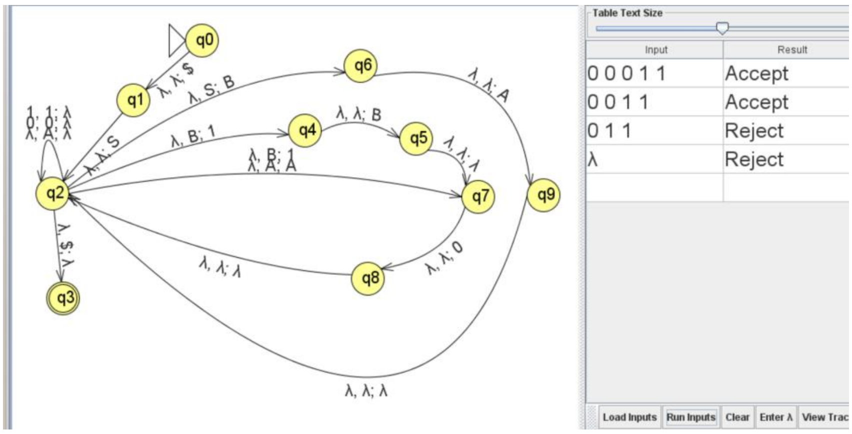
V = {S, A, B}

T = {0, 1}　　下面根据 G 设计 PDA

P: $\begin{cases} S \to AB \\ A \to 0A \mid \varepsilon \\ B \to 0B1 \mid 01 \end{cases}$

S: {S}





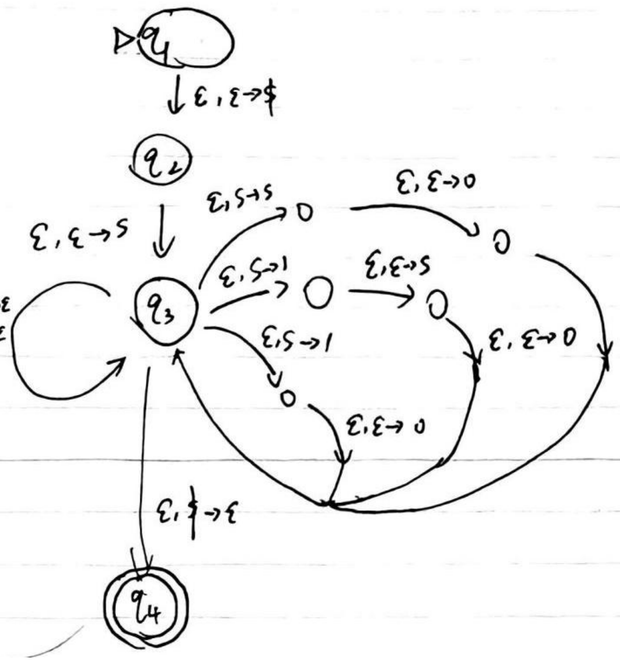
第二种方法如下：

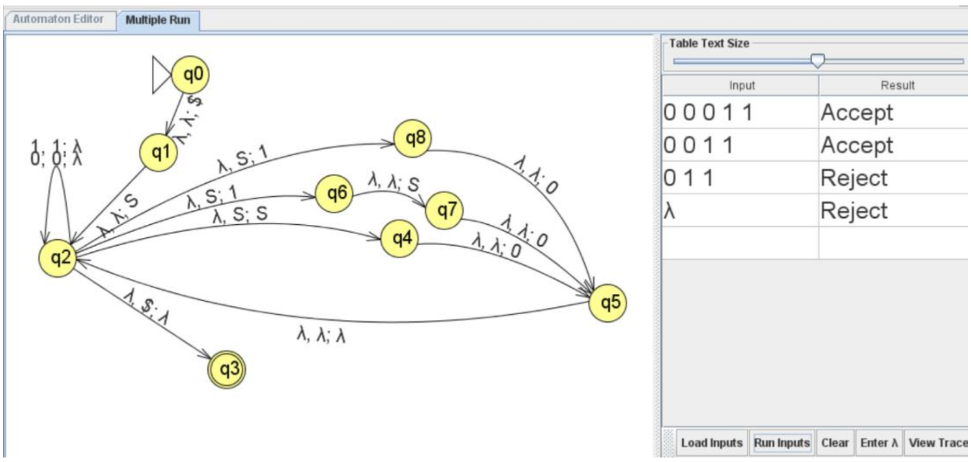
也可构造 $G=(V,T,P,S)$

$V=\{S\}$　$S=\{S\}$

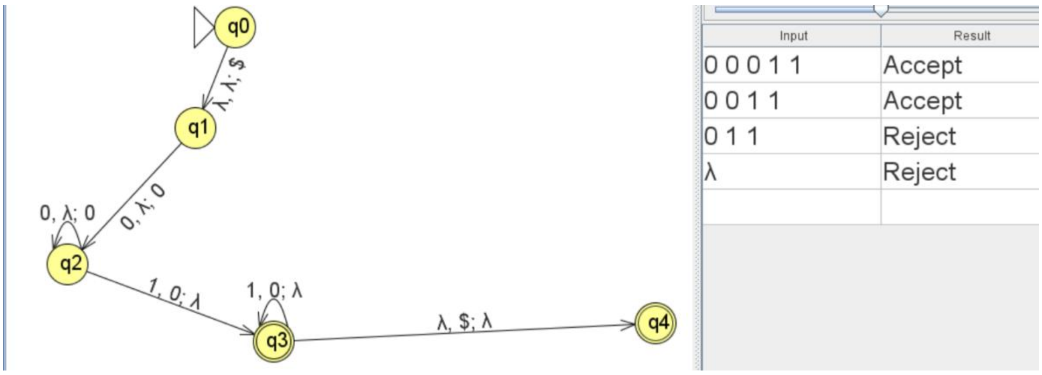
$T=\{0,1\}$

P: $S \to 0S1 \mid 0S \mid 01$　更简单吗?





第三种方法：直接做就行，但我更喜欢上述两个通用的方法：



2) $\{1^n0^n1^m0^m \mid n, m \geqslant 1\}$

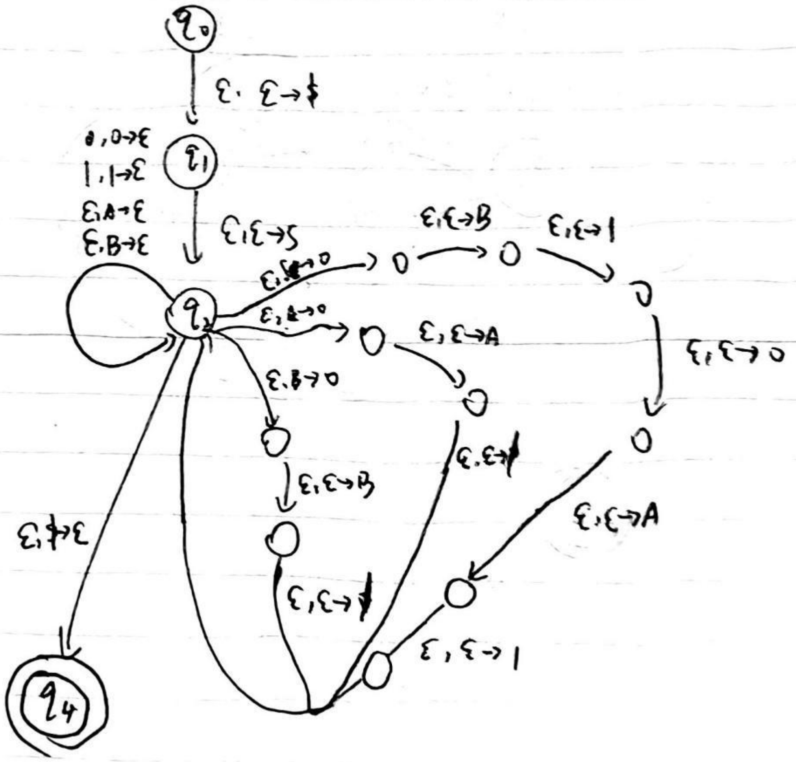
2) $\{1^n0^m1^n0^m \mid n,m \geq 1\}$

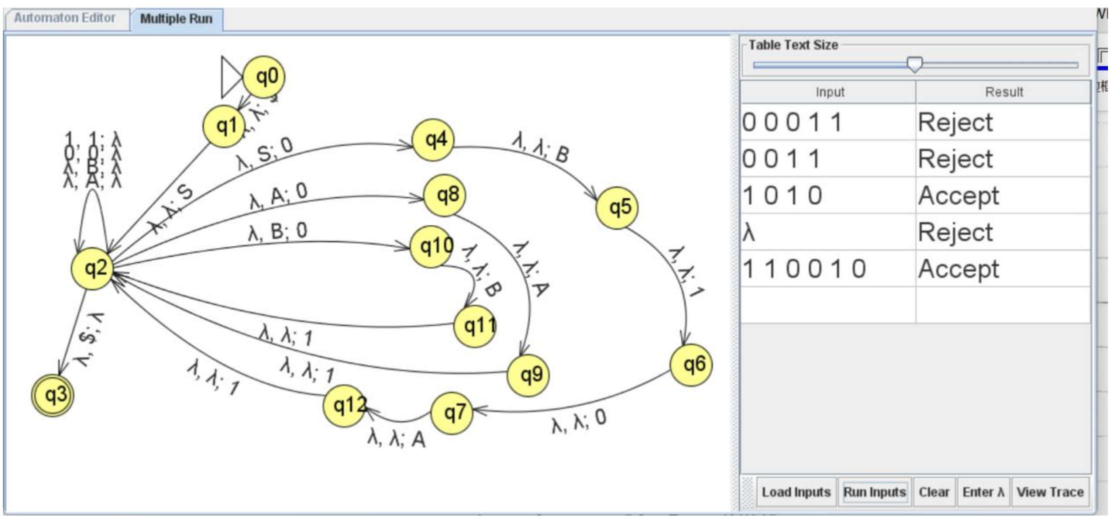
$G = (V, T, P, S)$

$V = \{S, A, B\}$

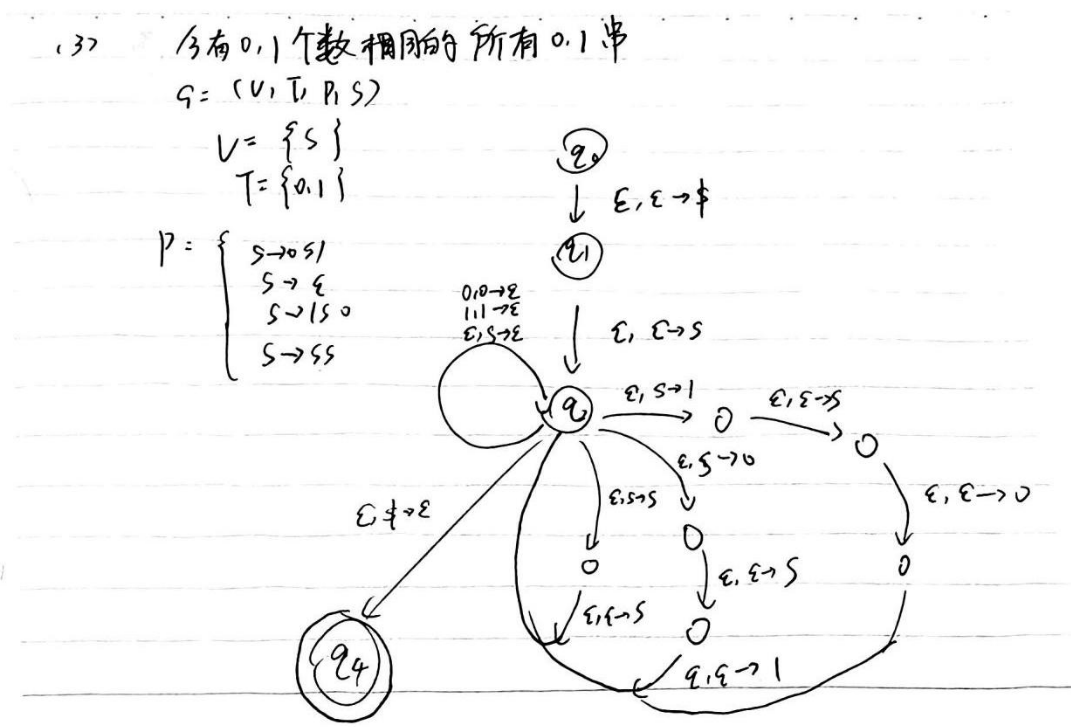
$T = \{0, 1\}$

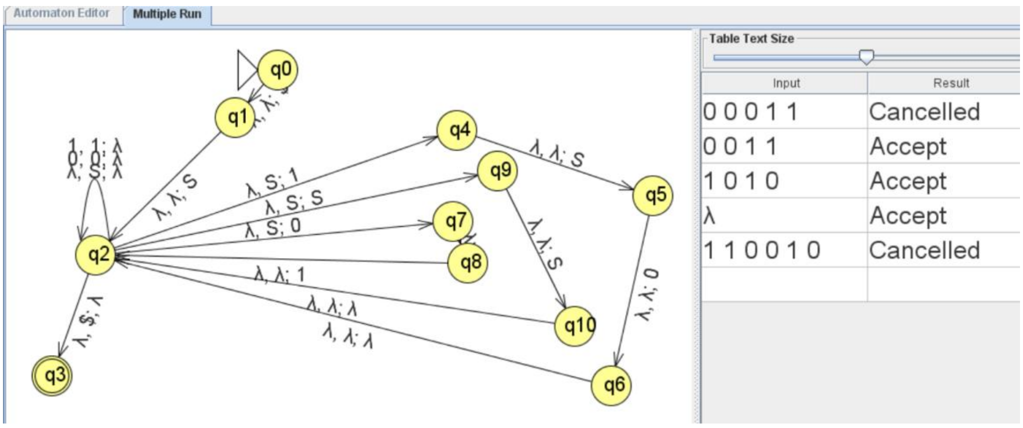
$P: \begin{cases} S \to 1A0 \mid 1B0 \\ A \to 1A0 \mid \varepsilon \\ B \to 1B0 \mid \varepsilon \end{cases}$



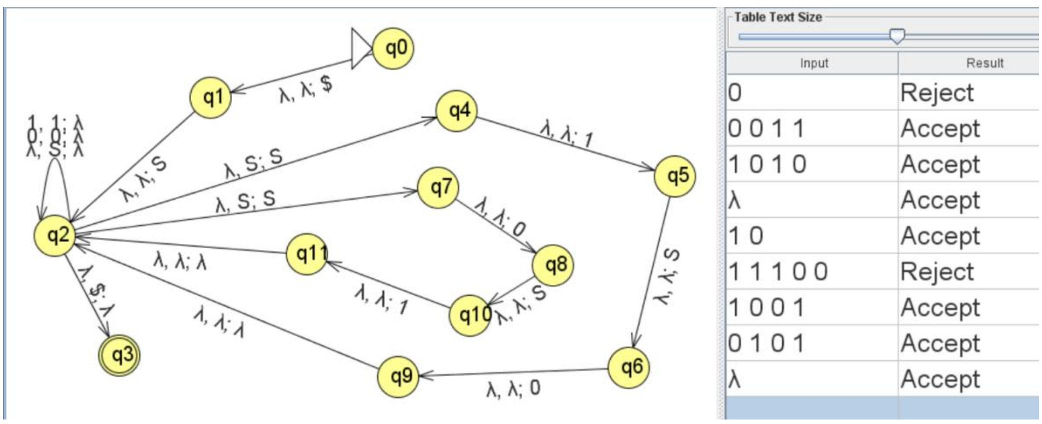


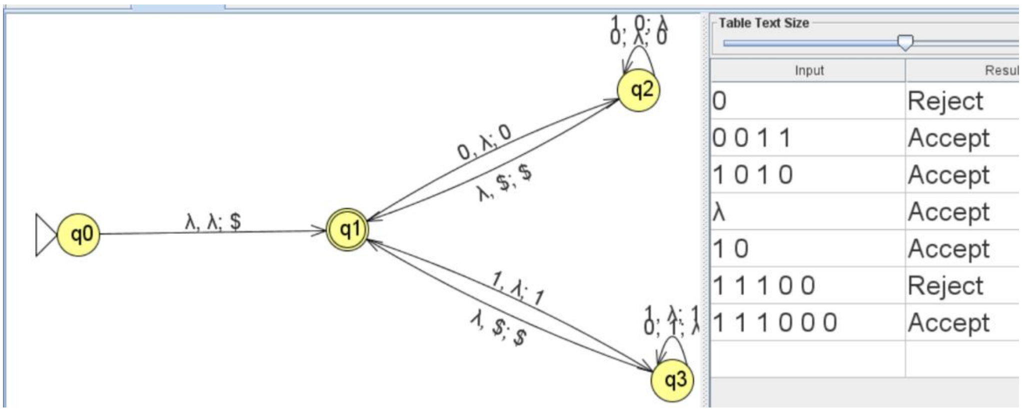
3） 含有 0 的个数和 1 的个数相同的所有 0，1 串





这里的拒绝状态出现了取消结果，关键原因是 S->SS 这个产生式会导致状态无限膨胀，进而使得拒绝态可以无限循环，这里给出下面的更加安全的 PDA：我们的产生式子就应该修改为：S→0S1S｜1S0S｜ε 再下面是直接做的方法，来 1（0）如果有 0（1）就抵消





Input	Result
0	Reject
0 0 1 1	Accept
1 0 1 0	Accept
λ	Accept
1 0	Accept
1 1 1 0 0	Reject
1 1 1 0 0 0	Accept

2. 构造一个 PDA，使它等价于下列文法：

S→aAA，A→aS|bS|a

$$\begin{cases} S \to aAA \\ A \to aS \mid bS \mid a \end{cases}$$