2152118 史君宝 形式语言与自动机 第五次作业

1. 设 T = {0, 1,(,), +, *, ϕ, 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。

解：我们先考虑一下有关{0,1}的正则表达式(0+1) $(0+1)^*$ $(0+1)^+$ (ε+1) ϕ
我见过将 ε 放在正则表达式中的，例如$(0+1+\varepsilon)^*$ 但是我没有见过$(\phi+1)^*$ 的
所以我会采用一个分流。

对于 CFG G = (V, T, P, S) = ({S,T}, {0,1,(,), +, *, ϕ, e }, P, S)，
其中非终结符就是 V = {S,T}
对于终结符来说 T = { 0, 1, (,), +, *, ϕ, e }
对于产生式来说 P ：S -> T | ϕ
T -> 0 | 1 | e | T+T | TT | T^* | T^+ | (T)
对于初始状态来说 就是 S

在上面由于我们有讲过ϕ与其他结合的，所以我利用了一个 T 将它分流出去了。

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。

证明 ：首先我们知道 CFG 中任何一个产生式的右边都不是 ε 。
也就是说叶子结点的个数一定等于 w 的长度，所以说叶子结点的个数为 n。

在这里我们采用数学归纳法证明：
当 m = 1 的时候，就是我刚才所说的 可以将 w = w1w2w3……wn
由于 m = 1 ，所以 S -> w1w2w3……wn
所以结点数就是 叶子结点 + 根节点 = n + 1

当 m > 1 的时候 S -> X1X2X3……Xk 我们也可以将 w = w1w2w3……wk
对于 1<=i<=k 的时候，当 Xi 是终结符的时候，那么 Xi = wi 结点数 wi
当 Xi 不是终结符的时候，我们可以知道 Xi => wi 即可以推导出来。
此时它的推导过程中产生的结点为 wi + mi （包括 Xi 这个结点）

所有推导结束的时候 结点数为 1+w1+w2+……+wk+m1+m2+……+mk
其中 w1+w2+……+wk = n m1+m2+……+mk = m-1 （S 已经推导了一次了）
上述结点数是 m+n 得证。

3. 假设在上一题中除了 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 个结点不可能更多。

证明 ： 在这里我们采用数学归纳法证明：
当 m = 1 的时候，就是我刚才所说的 可以将 w = w1w2w3……wn
由于 m = 1 ，所以 S -> w1w2w3……wn
由于右端为 ε 的话，那么就不能满足 w 不是 ε 的条件，所以上述不可能出现 ε
所以结点数就是 叶子结点 + 根节点 = n + 1
此时 m = 1 ， n + 2m -1 = n + 1

当 m > 1 的时候 S -> X1X2X3……Xk 我们也可以将 w = w1w2w3……wk
对于 1<=i<=k 的时候，当 Xi 是终结符的时候，那么 Xi = wi 结点数 wi
当 Xi 不是终结符的时候，我们可以知道 Xi => wi 即可以推导出来。
此时它的推导过程中推导次数是 mi，产生 wi 的数 （包括 Xi 这个结点）
我们根据题目有所有的 Xi 下面的结点数最多是 wi + 2mi -1

所有推导结束的时候 结点数为 1+w1+w2+……+wk+2m1+2m2+……+2mk -i
其中 i 是所有非终结符 Xi 的个数
其中 w1+w2+……+wk = n m1+m2+……+mk = m-1 （S 已经推导了一次了）
所以上述 1+w1+w2+……+wk+2m1+2m2+……+2mk -i = n + 2m -1 -i
由上可知 0<=i<=k 所以说结点数可能到达 n + 2m -1 但是不可能更多。
得证。

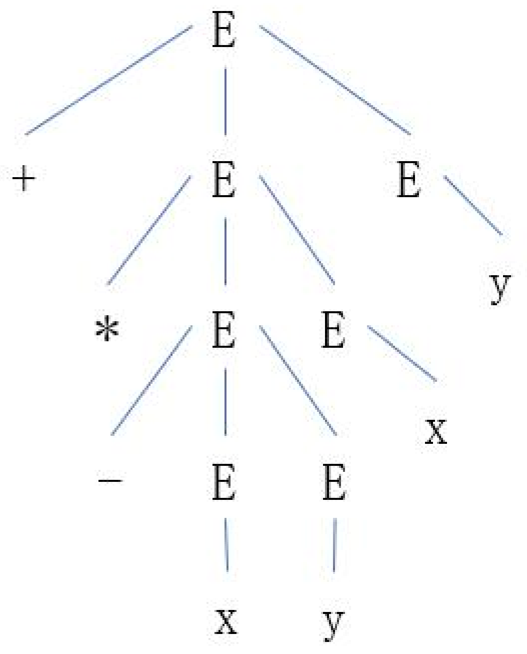
4. 下面的文法生成的是具有 xx 和 yy 操作数、二元运算符 +、-和*的前缀表达式：
E -> +EE | *EE | -EE | x | y
(a)找到串 +*-xyxy 的最左推导、最右推导和一棵语法分析树。
(b)证明这个文法是无歧义的。

（a）

最左推导：
解：E => +EE
=> +*EEE
=> +*-EEEE
=> +*-xEEE
=> +*-xyEE
=> +*-xyxE
=> +*-xyxy

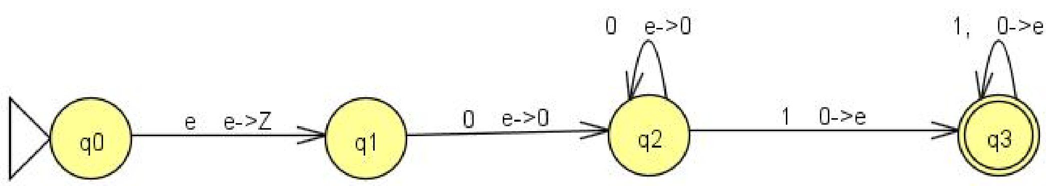
最右推导：
E => +EE
=> +Ey
=> +*EEy
=> +*Exy
=> +*-EExy
=> +*-Eyxy
=> +*-xyxy

下面我们给出相应的语法分析树：



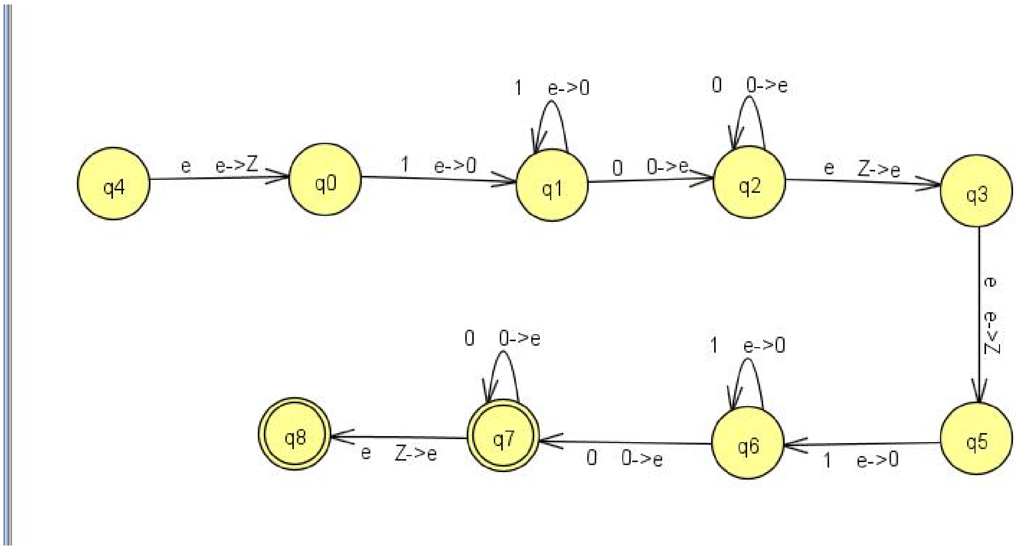
（b）
解：我们仔细观察上述文法的产生式：
E -> +EE | *EE | -EE | x | y
我们发现，尽管前三个分别是 +EE *EE -EE 的，有 EE 的相同，但+，-，*不同
后面两个是 x，y 也是不相同的。
总的来说不论最终产生的语言语句是什么样的，对于语句中的每一个符号，比如
+，-，*，x，y 来说都有固定的产生式：比如+，就是 E -> +EE
所以说不管是正面的推导还是反面的还原，每一个终结符都是有着固定的产生式，
则它的语法分析树是固定的，即文法是无歧义的。
得证。

5. 对于下面语言，分别构造出接受它们的 PDA
（1）{0^n1^m，n>=m>=1}

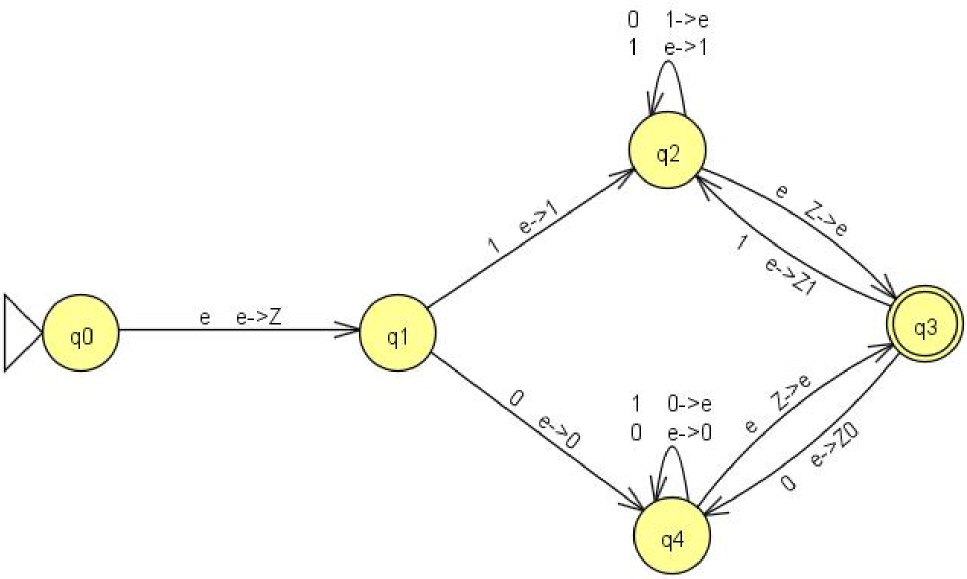


其中 e 是 ε

（2）$\{1^n0^n1^m0^m，n>=1\ m>=1\}$



（3）含有 0 的个数和 1 的个数相同的所有 0，1 串



6．构造一个 PDA，使它等价于下列文法：
S→aAA，A→aS|bS|a