

# Compilers Assignment: 正则表达式

10185101210 陈俊潼

## 1. 描述下列正规式（正则）定义的语言

a)  $0(0 \mid 1)^*0$

以 0 开头，0 结尾的只含 0 和 1 的字符串。

b)  $((e \mid 0)1^*)^*$

任意个 e 或 0 后伴随任意个 1 的循环。

c)  $(0 \mid 1)^*0(0 \mid 1)(0 \mid 1)$

以 000、001、010 或 011 结尾的任意长度大于等于 3 的只含 0 和 1 的字符串。

d)  $0^*10^*10^*10^*$

有且只有 3 个 1 的任意只含 0 和 1 的字符串。

e)  $(00 \mid 11)^*((01 \mid 10)(00 \mid 11)^*(01 \mid 10)(00 \mid 11)^*)^*$

该字符串只含有 1 和 0，且满足：

1、总长度为偶数

2、两两组合，大多数情况下应该为两个 0 或者 两个 1，允许出现例外（即一个 0 一个 1），且例外的对数也为偶数个。

## 2. 为下列语言写出正规定义(或正规式、正规文法、有限自动机)

a) 包含五个元音，且按顺序排列的所有字母串（RE）

$aa*ee*ii*oo*uu*$

b) 字母按字典升序排列的所有字母串（RE）

$a*b*c*d*e*f*g*h*i*j*k*l*m*n*o*p*q*r*s*t*u*v*w*x*y*z*$

c) 以 /\* 开始，\*/ 结束的注释，中间不能包含 /\*，除非包含在 " 和 " 中（RE）

$/\*((P\backslash**)\mid(P/*))*("Q*")*((P\backslash**)\mid(P/*))*\backslash*/$

其中含有两个正则定义 P, Q:

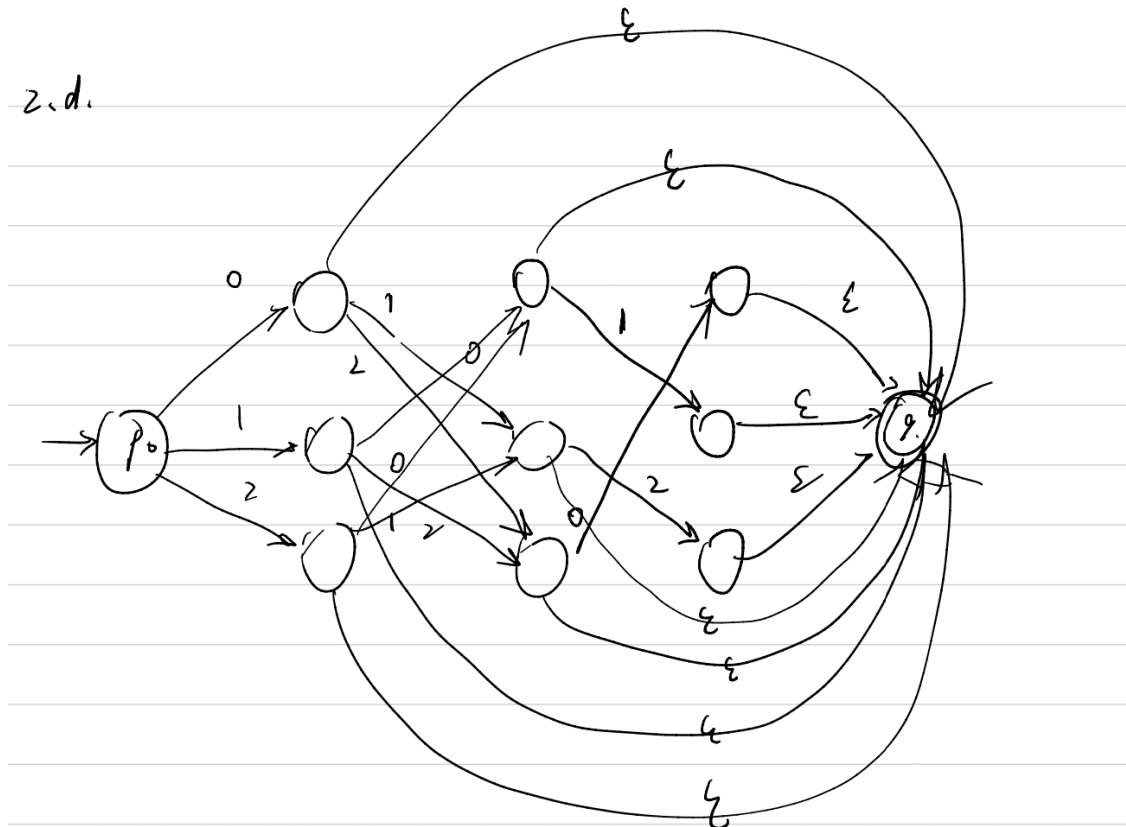
$P \rightarrow$  任意除了  $*$  的字符

$Q \rightarrow$  任意除了  $''$  的字符

使用  $\backslash *$  来表示匹配星号来和一元运算符  $*$  做区分。

d) 不包含重复数字的数字串 (正规定义)

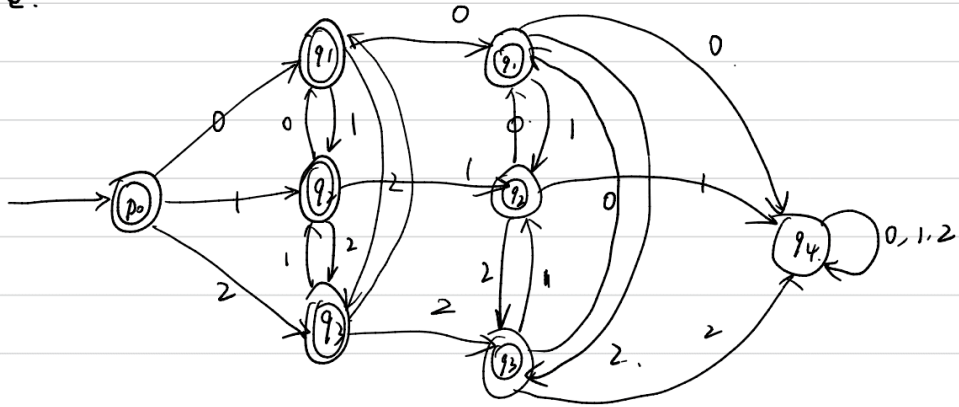
构造 DFA 如下。由于 9 个数字的自动机过于繁杂, 这里绘制的是只含有 0, 1, 2 三个数字的情况, 更多的数字只是在中间有更多的状态和转移。



e) 包含至多一个重复数字的数字串 (正规定义)

构造 DFA 如下。由于 9 个数字的自动机过于繁杂, 这里绘制的是只含有 0, 1, 2 三个数字的情况, 更多的数字只是在中间有更多的状态和转移。

2. e.



f) 包含偶数个0和奇数个1的0、1串（正规定义）

2. f.

