计算理论导论

习题二: 正则表达式、泵引理

中国人民大学　信息学院　崔冠宇　2018202147

1. Design REs: **1.18** (b)(e)(l)(n).

(b) $\{w|\ w$ contains at least three 1s$\}$,

(e) $\{w|\ w$ starts with 0 and has odd length, or starts with 1 and has even length$\}$,

(l) $\{w|\ w$ contains an even number of 0s, or contains exactly two 1s$\}$,

(n) All strings except the empty string.

解:

(b) 思路: 三个 1 之间可以有任意字符。于是可得 $(0+1)^*1(0+1)^*1(0+1)^*1(0+1)^*$，或 $\Sigma^*\,1\,\Sigma^*\,1\,\Sigma^*\,1\,\Sigma^*$。

(e) 思路: 先根据要求写出两个语言，再用 + 并起来。于是可得 $0((0+1)(0+1))^* + 1((0+1)(0+1))^*(0+1)$，或 $0(\Sigma\Sigma)^*+1(\Sigma\Sigma)^*\Sigma$。

(l) 思路: 对于前半部分，可以先设计一个 DFA，再将其转化为 RE；对于后半部分，可以直接写出。于是可得 $(1+01^*0)^*+0^*10^*10^*$。

(n) 思路：可以直接写出。于是可得 $(0+1)(0+1)^*$，或 $\Sigma\Sigma^*$。

2. RE $\to$ NFA and NFA $\to$ DFA: **1.17**

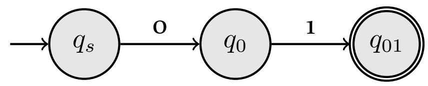
a. Give an NFA recognizing the language $(01 \cup 001 \cup 010)^*$.

b. Convert this NFA to an equivalent DFA. Give only the portion of the DFA that is reachable from the start state.

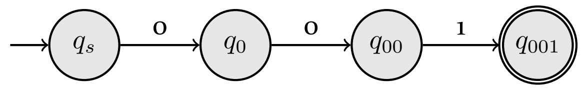
解:

a. 先分别设计出接受 01、001、010 的 NFA，然后将其组装成识别 $(01 \cup 001 \cup 010)$ 的 NFA，再将其构造成识别 $(01 \cup 001 \cup 010)^*$ 的 NFA，最后将其化简。

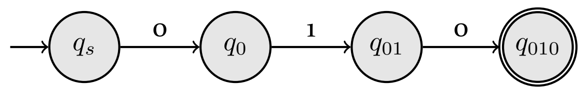
(1) 识别 01 的 NFA:



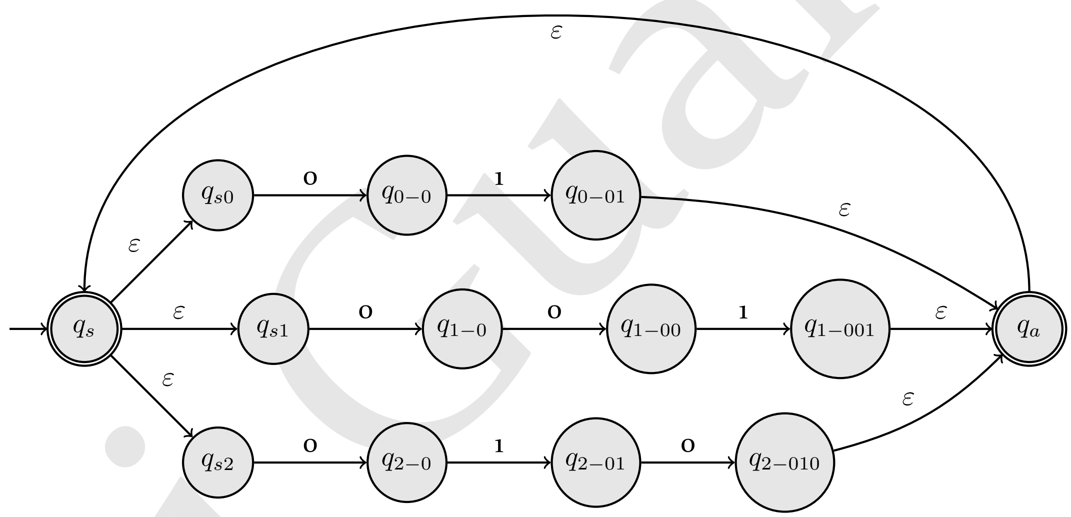
(2) 识别 001 的 NFA:



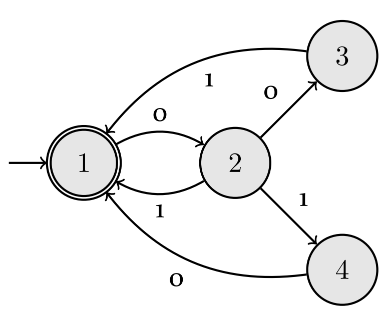
(3) 识别 010 的 NFA:



(4) 识别 $(01 \cup 001 \cup 010)^*$ 的 NFA:



化简为一个不含 ε 边的四状态 NFA:



b. 运用 **Lemma 1.19** 将其转换为等价的 DFA $A = (Q', \Sigma, \delta', q_0', F')$。其中

$Q' = \mathcal{P}(Q) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}, \{1,2,3\}, \{1,2,4\},$

$\{1,3,4\}, \{2,3,4\}, \{1,2,3,4\}\}$,

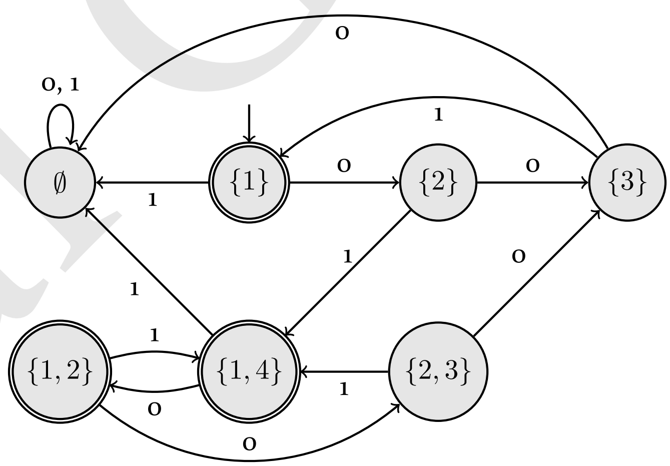
$q_0' = \{q_0\} = \{1\}$,

$F' = \{q \in Q' | q \cap F \neq \emptyset\} = \{\{1\}, \{1,2\}, \{1,3\}, \{1,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{1,2,3,4\}\}$,

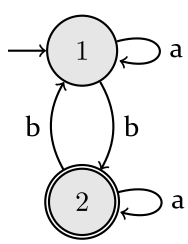
$\delta'(R, a) = \bigcup_{q \in R} E(\delta(q, a))$ 由下表给出:

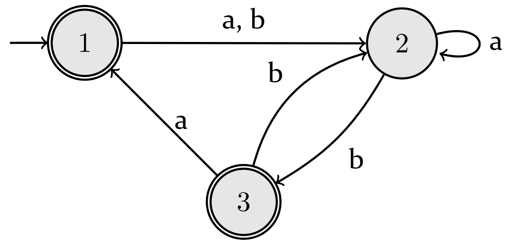
	0	1
$\emptyset$	$\emptyset$	$\emptyset$
$\{1\}$	$\{2\}$	$\emptyset$
$\{2\}$	$\{3\}$	$\{1,4\}$
$\{3\}$	$\emptyset$	$\{1\}$
$\{4\}$	$\{1\}$	$\emptyset$
$\{1,2\}$	$\{2,3\}$	$\{1,4\}$
$\{1,3\}$	$\{2\}$	$\{1\}$
$\{1,4\}$	$\{1,2\}$	$\emptyset$
$\{2,3\}$	$\{3\}$	$\{1,4\}$
$\{2,4\}$	$\{1,3\}$	$\{1,4\}$
$\{3,4\}$	$\{1\}$	$\{1\}$
$\{1,2,3\}$	$\{2,3\}$	$\{1,4\}$
$\{1,2,4\}$	$\{1,2,3\}$	$\{1,4\}$
$\{1,3,4\}$	$\{1,2\}$	$\{1\}$
$\{2,3,4\}$	$\{1,3\}$	$\{1,4\}$
$\{1,2,3,4\}$	$\{1,2,3\}$	$\{1,4\}$

在上表中从 $\{1\}$ 开始进行遍历（用红色表示），发现仅有 $\emptyset$、$\{1\}$、$\{2\}$、$\{3\}$、$\{1,2\}$、$\{1,4\}$、$\{2,3\}$ 可被遍历到，绘制对应的 DFA 如下图:



3. DFA $\rightarrow$ RE: **1.21** Use the procedure described in **Lemma 1.60** to convert the following finite automata to regular expressions.

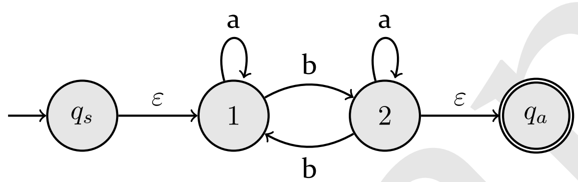




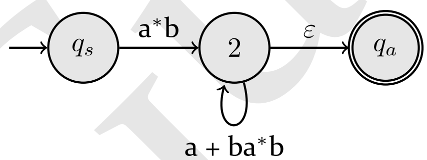
解:

(1)

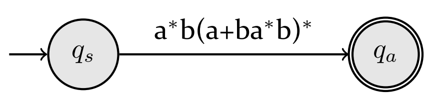
1. 增加新的起始、接受状态:



2. 去掉状态 1:



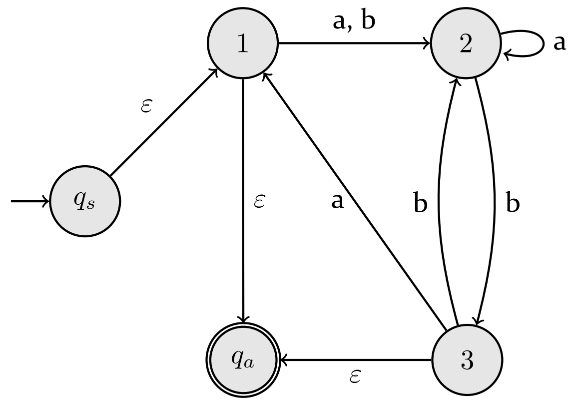
3. 去掉状态 2:



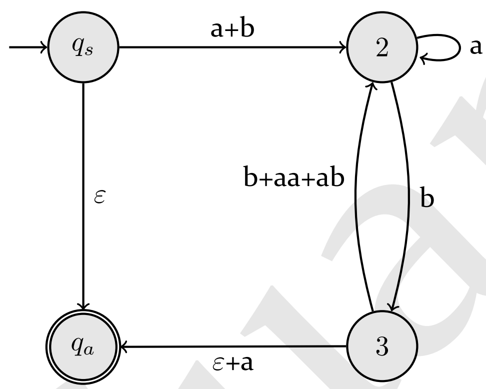
于是最终结果为 a*b(a+ba*b)*。

(2)

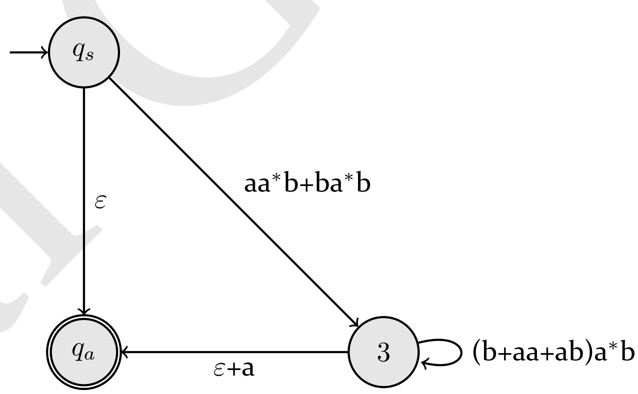
1. 增加新的起始、接受状态:



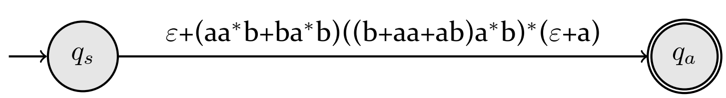
2. 去掉状态 1:



3. 去掉状态 2:



4. 去掉状态 3:



于是最终结果为 ε+(aa*b+ba*b)((b+aa+ab)a*b)*(ε+a)。

4. Let L be a regular language. Let $L_0 \subseteq L$. Is L_0 necessarily regular? Why?

解: L_0 不一定是正则的，反例如下:

设 $\Sigma = \{0, 1\}$，取 $L = (0+1)^* = \Sigma^*$，$L_0 = \{0^n 1^n | n \geq 0\}$。因为 L 能写成正则表达式，所以 L 是正则语言；同时容易看出 $L_0 \subseteq L$，下面证明 L_0 不是正则的。

若不然，假设 L_0 为正则语言，设其由泵引理得到的泵长度为 p。考察串 $w = 0^p 1^p$，由于 $|w| \geq p$，利用泵引理得到 w 可以写成 xyz 三部分，其中 $|y| > 0$ 且 $|xy| \leq p$。由于 $|xy| \leq p$，因此 y 只可能具有 $0^k (k > 0)$ 的形式，于是 $xyyz = 0^{p+k} 1^p \notin L_0$，矛盾，即 L_0 不是正则语言。

5. A language L is called finite if it contains finitely many strings. Prove that every finite language is regular.

证明思路: 考虑到正则语言与正则表达式的等价性，用构造有限语言的正则表达式的方法来证明。证明分两步，先证明任意单个（有限长）字符串 w 是正则表达式；再应用加法规则证明结论。

证明: 设有限语言 $L = \{w_i | 1 \leq i \leq n\}$。

1. 先证明任意有穷字符串 $w = a_1 a_2 \cdots a_m$ 是正则表达式。因为每个单字符都是正则表达式，反复应用**有限次**正则表达式的连接规则，可得 w 是正则表达式。

2. 在证明 L 可用正则表达式表出。由上一步证明，各 w_i 是正则表达式，反复应用**有限次**加法规则可得 $\sum_{i=1}^{n} w_i = w_1 + w_2 + \cdots + w_n$ 是正则表达式。

因为 L 有对应的正则表达式，根据正则表达式与正则语言的等价性，有限语言 L 是正则语言。$\square$

6. Use the Pumping Lemma to prove that the following languages over $\Sigma = \{0, 1\}$ are not regular.

(a) The language L_1 consisting of all palindromes. A palindrome is a string that equals its own reverse, such as 00100 or 110011.

(b) The language L_2 consisting of all strings in which the number of 1's is exactly three times the number of 0's.

(c) The language $L_3 = \{www | w \in \{0, 1\}^*\}$.

解: 假设上述各 L_i 为正则语言，设 p 是泵引理给出的泵长度。

(a) 取 $w = 0^p110^p$。若 w 能被写成 xyz 的形式，其中 $|y| > 0$，$|xy| \le p$，则 y 一定具有 $0^k(k > 0)$ 的形式，于是 $xyyz = 0^{p+k}110^p \notin L_1$，矛盾。

(b) 取 $w = 0^p1^{3p}$。类似于上面的讨论，y 一定具有 0^k 的形式，于是 $xyyz = 0^{p+k}1^{3p} \notin L_2$，矛盾。

(c) 取 $w = 0^p10^p10^p1$。类似于上面的讨论，y 一定具有 0^k 的形式，于是 $xyyz = 0^{p+k}10^p10^p1 \notin L_3$，矛盾。