理论计算机科学基础

第一章习题课

晏荣杰

中科院软件所

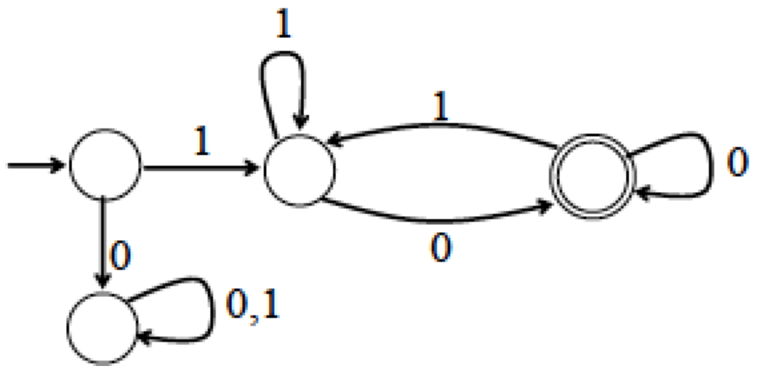
计算机科学国家重点实验室

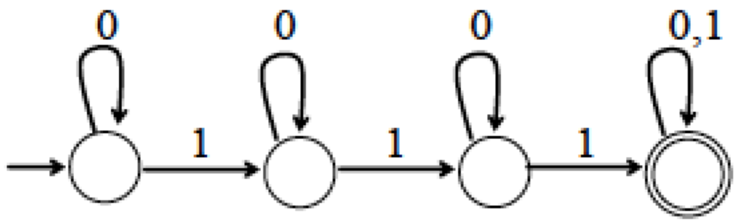
第 1 章正则语言, 小结

- 主要概念 (是如何定义的?)
  正则语言，DFA, NFA, 正则表达式。1.6
- 主要结论 (是如何证明的?)
  正则语言类 = DFA 接受的语言类 = NFA 接受的语言类 = 正则表达式表示的语言类
- 主要转换算法
  NFA2DFA, RE2NFA，NFA2RE 1.16，1.21，1.28
- 正则语言关于并、交、补、连接及星的封闭性 1.14
  用 DFA 构造正则语言的并、交、补
  用 NFA 构造正则语言的并、连接和 *
- 归纳定义和归纳证明的方法
  如 RE 的定义、RE 表示的语言、以及 RE2NFA 的转换过程
- 泵引理、Myhill-Nerode 定理，以及它们的简单应用 (证明非正则性) 1.29，1.30, 1.48

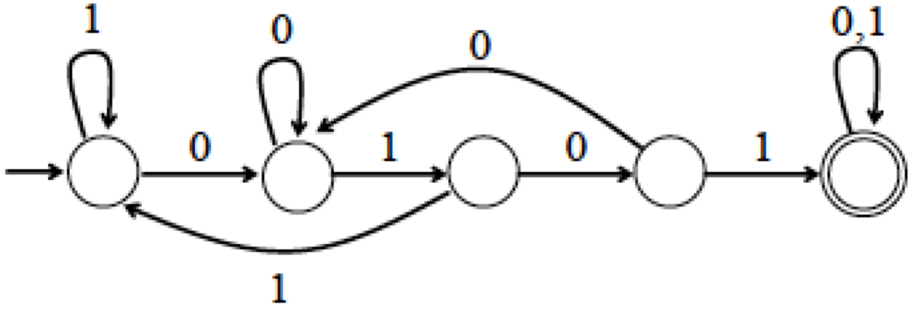
第一章 正则语言

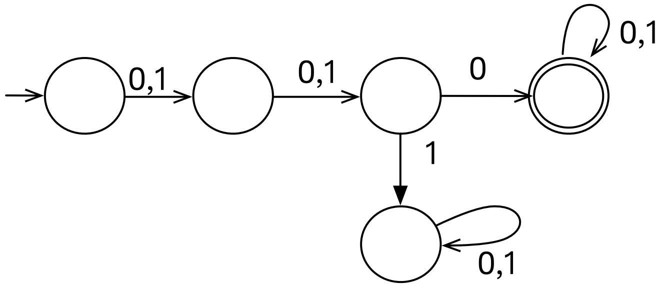
- 1.6 画出识别下述语言的DFA的状态图，字母表均为{0,1}
- a){w | w从1开始以0结束}　　　b){w | w至少有3个1}





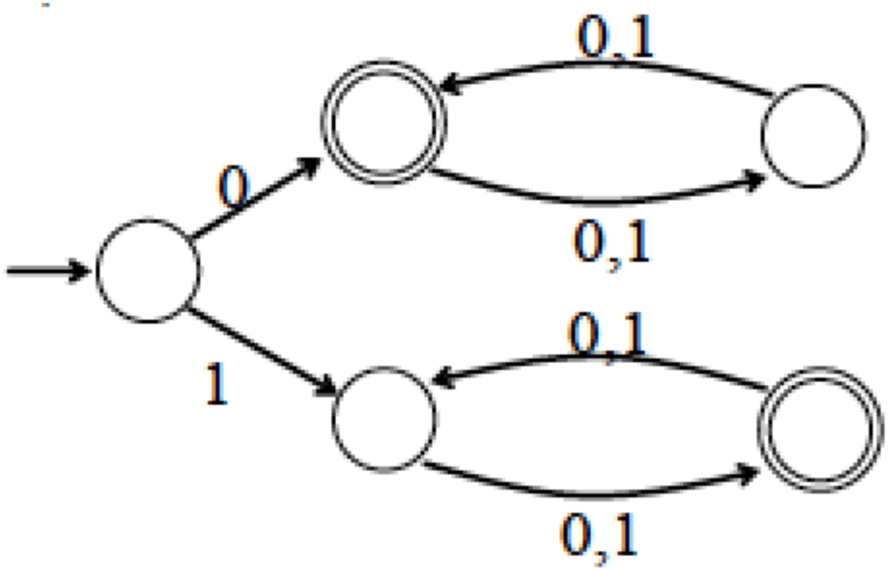
- c) {w | w含有子串0101}　d) {w | w的长度不小于3，且第三个符号为0}



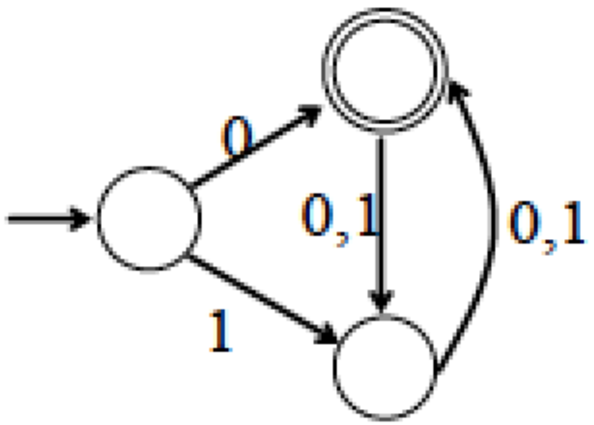


第一章 正则语言

- 1.6 画出识别下述语言的DFA的状态图，字母表均为{0,1}
- e) {w | w从0开始且为奇长度，或从1开始且为偶长度}



或



第一章 正则语言

1.14

- a 证明：设M是一台识别语言B的DFA，交换M的接受状态与非接受状态得到一台新的DFA，则这台新的DFA是识别语言B 的补集。因而，正则语言类在补运算下封闭。

解：M是DFA, 表示为$M=(Q,\Sigma,\delta,q_0,F)$,

令交换接受状态与非接受状态后的DFA为$N=(Q,\Sigma,\delta,q_0,Q-F)$,

设$\omega=\omega_1\omega_2\ldots\omega_n$是字母表上任意字符串，

因为M与N均为DFA,所以必然存在Q中状态序列$r_0,r_1,\ldots,r_n$,使得：

$r_0=q_0$, $\delta(r_i, \omega_{i+1})=r_{i+1}$, $i=0,\ldots,n-1$

1)若$r_n \in F$，则$r_n \notin Q-F$, 有M识别ω，N拒绝ω;

2)若$r_n \notin F$,则$r_n \in Q-F$,即N识别ω，M拒绝ω,

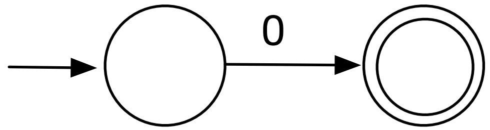
所以N接受B的补集，即B的补集正则。

因此，正则语言类在补运算下封闭。

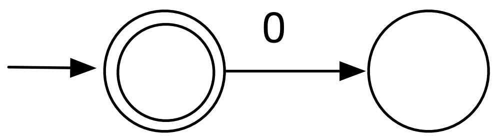
第一章 正则语言

1.14

- b 举例说明：设M是一台识别语言C的NFA，交换M的接受状态与非接受状态得到一台新的NFA，这台新的NFA不一定识别C的补集。NFA识别的语言类在补集下封闭吗？解释你的回答。



识别{0}



识别{ϵ}

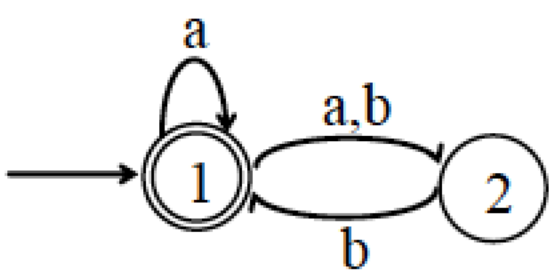
交换M的接受状态与非接受状态得到的新的NFA不一定识别C的补集。

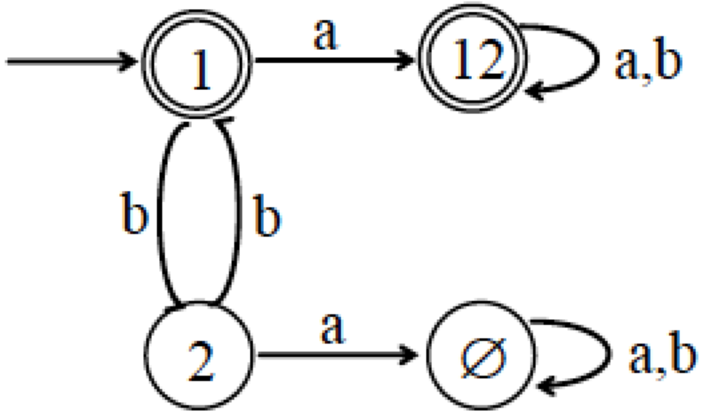
由于NFA识别的语言类与DFA识别的语言类相同，即正则语言类。由a的证明，正则语言类在补运算封闭。可知，**NFA识别的语言类---正则语言类在补运算下封闭**。

若 NFA 识别语言A，必有等价的DFA识别A，从而由a知，可以交换DFA的接受与非接受状态，构造识别A的补集的DFA，则必有等价的NFA识别A的补集。**只是，该 NFA 未必由原来的NFA交换接受与非接受状态构造而成。**

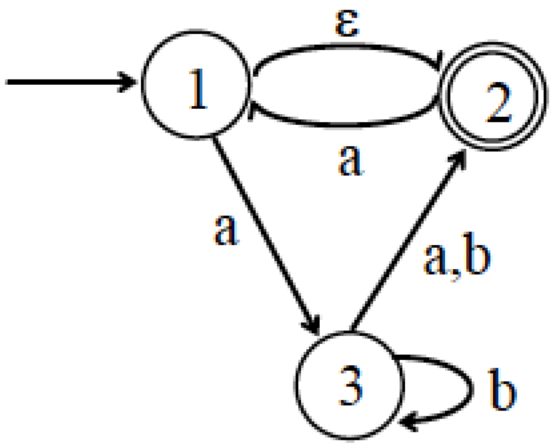
第一章 正则语言

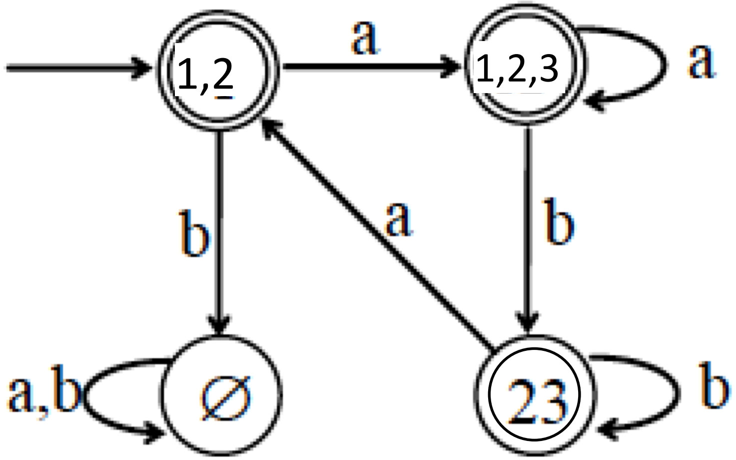
- 1.16 使用定理1.19给出的构造，把下图的两台非确定型有穷自动机转换成等价的确定型有穷自动机
- （a）





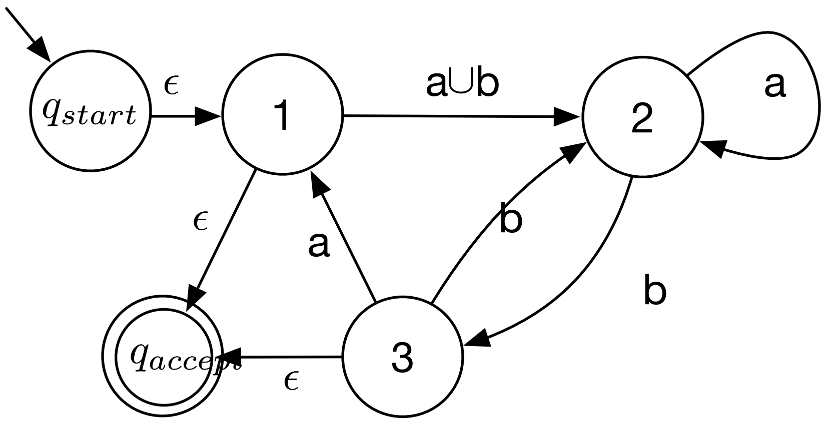
（b）



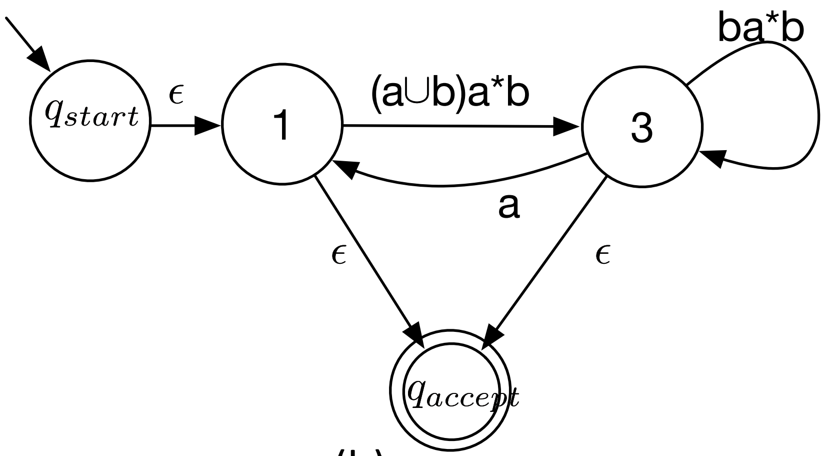


第一章 正则语言

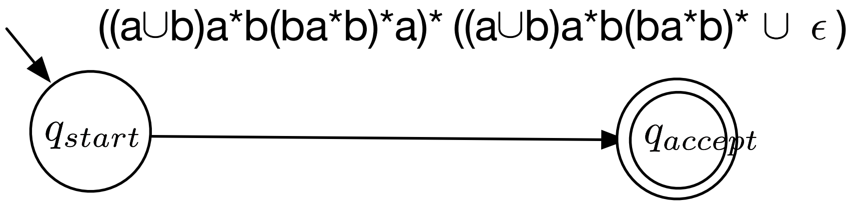
1.21使用引理1.32中描述的过程，把下图的有穷自动机转换成正则表达式：消去方式：2、3、1



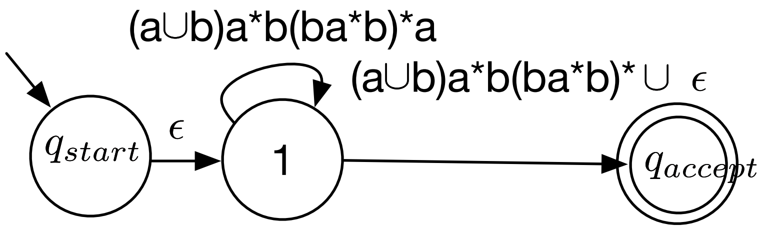
(a)



(b)



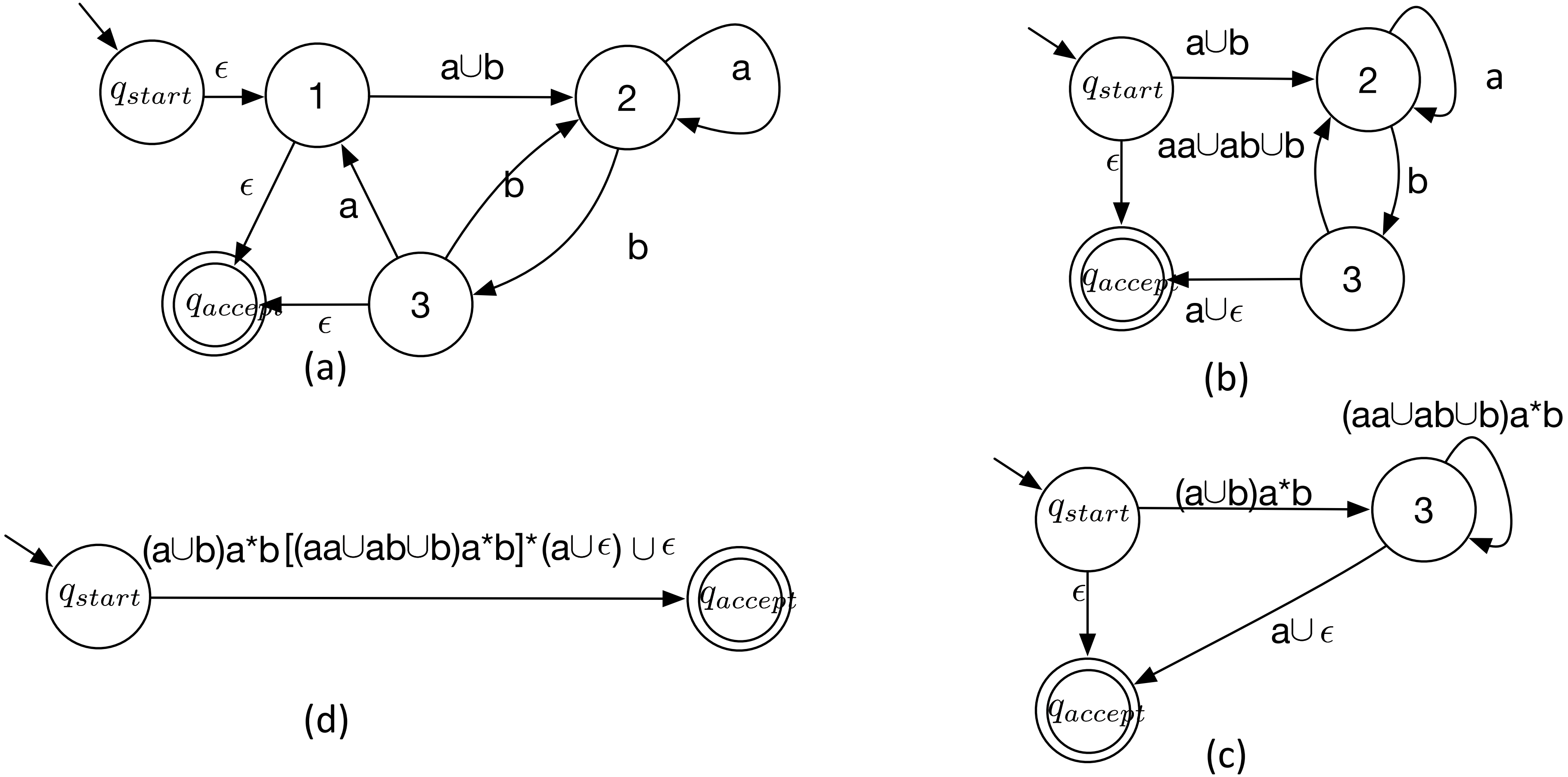
(d)



(c)

第一章 正则语言

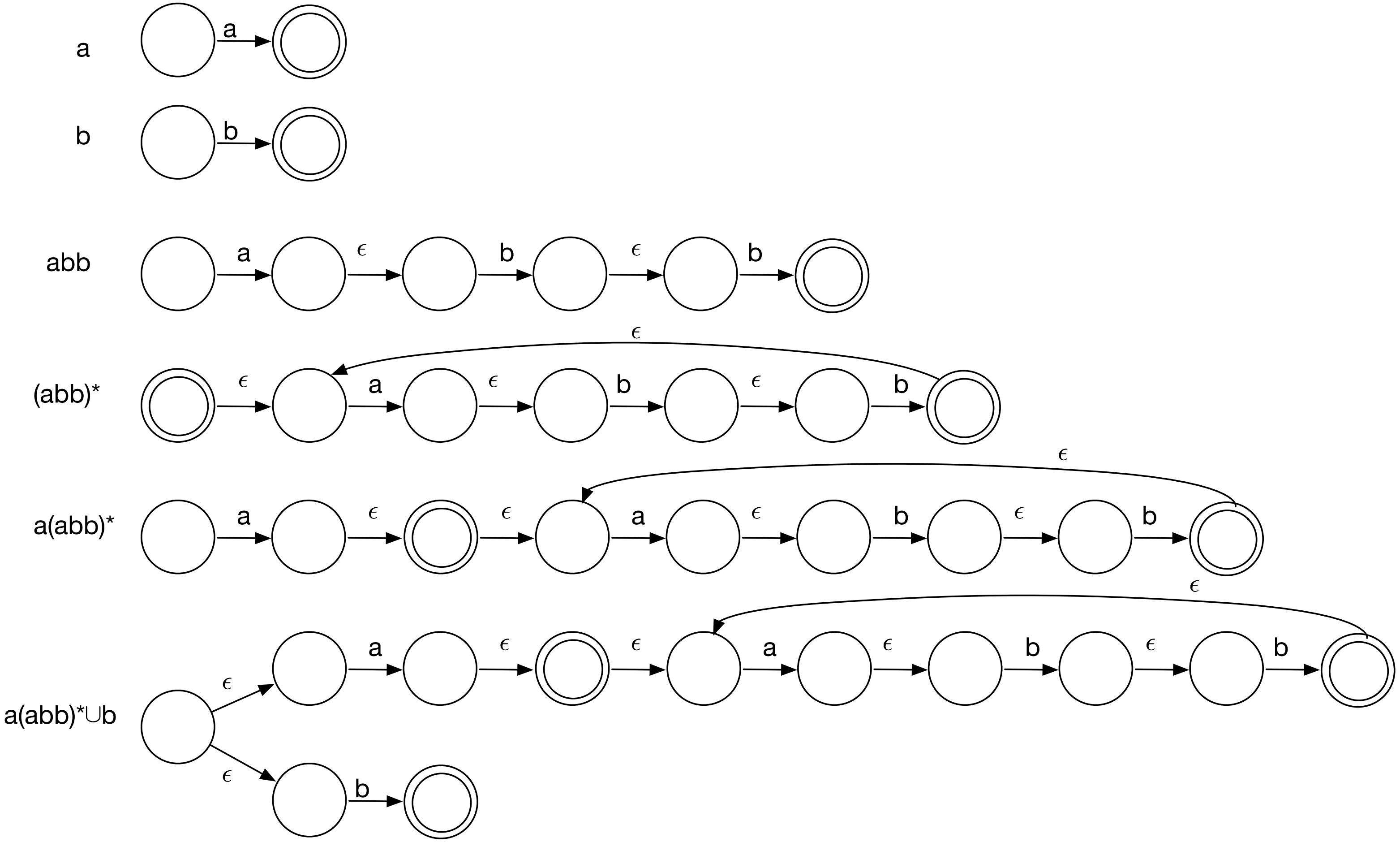
1.21使用引理1.32中描述的过程，把下图的有穷自动机转换成正则表达式。消去方式：1、2、3



第一章 正则语言

1.28 使用定理1.28给出的过程将下述正则表达式转换成NFA

- a. a(abb)*∪ b



第一章 正则语言

1.29利用泵引理证明下述语言不是正则的。

- b. $A_2=\{www \mid w\in\{a,b\}^*\}$.

证明：假设A_2是正则的。设p是泵引理给出的关于A_2的泵长度。

令$s=a^pba^pba^pb$,

∵s是A_2的一个成员，且s的长度大于p，所以泵引理保证s可被分成3段s=xyz且满足泵引理的条件。

根据条件2，y中只含a，所以xyyz中第一个a的个数将比后两个a的个数多，故xyyz不是A_2的成员。违反泵引理的条件3，矛盾。

∴A_2不是正则的。

第一章 正则语言

- 1.30 指出下述关于0*1*不是正则语言的“证明”的错误(由于0*1*是正则语言，所以“证明”存在错误):
  - 用反证法证明。假设0*1*是正则的。设p是泵引理给出的关于0^*1^*的泵长度。选择s为字符串0^p1^p。已知s是0^*1^*的一个成员，但例1.38已证明s不能被抽取。这样得到矛盾，因此0^*1^*不是正则的。

解：在例1.38中的语言是$\{0^n1^n \mid n\geq 0\}$，取s为字符串0^p1^p，s确实不能被抽取；但针对语言0^*1^*，s是能被抽取的。

将s分成三段s=xyz，由泵引理的条件2，y仅包含0，而xy^iz属于语言0^*1^*，即s能被抽取，故题设中的“证明”不正确。

第一章 正则语言

- 1.48. Myhill-Nerode定理。参见问题1.47， 设L是一个语言，X是一个字符串集合。如果X中的任意两个不同的字符串都是用L可区分的，则称X是用L两两可区分的。定义L的指数为用L两两可区分的集合中的元素个数的最大值。L的指数可能是有穷的或无穷的。
- b.证明：如果L的指数是一个有穷数k，则它被一台有k个状态的DFA识别。
- 令DFA M=$(Q, \Sigma, \delta, q_0, F)$, 其中
  - $Q = \{[x]_L | x \in \Sigma^*\}$.
  - $\delta([x]_L, a) = [xa]_L$
  - q_0= $[\varepsilon]_L$
  - F $= \{[x]_L | x \in L\}$.

对于L中的任意字符串s= $s_1 \dots s_n$, 在M中的状态序列为$r_0 \dots r_n$， 满足r_0 =q_0= $[\varepsilon]_L$，
$r_{i+1} = \delta(r_i, s_{i+1}) = [s_1 \dots s_{i+1}]_L$,由于s ∈ Σ^*,有$r_{i+1} = [s_1 \dots s_{i+1}]_L$ ∈Q。
$r_n = [s]_L$∈ $\{[x]_L | x \in L\}$=F。
所以M识别L。