

3-将NFA转换为正则表达式/正规式

原理

hompson's Construction Theorem（汤普森构造定理）具体内容如下：

1. 对于每个正则表达式R，都存在一个对应的NFA N，使得N接受与R匹配的所有字符串，并且不接受与R不匹配的字符串。
2. 对于每个NFA N，都存在一个对应的正则表达式R，使得R匹配与N接受相同的所有字符串。

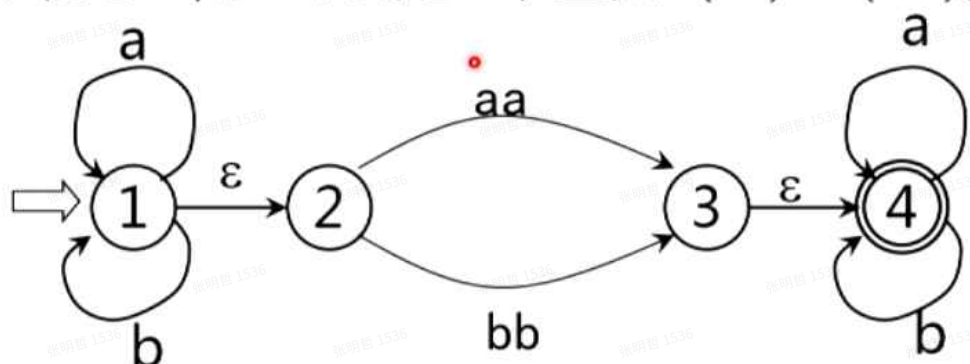
这个定理的主要思想是通过将正则表达式的各个部分映射到NFA的状态和转移来构建对应的NFA，或者通过对NFA的状态和转移进行一系列操作来构建对应的正则表达式。具体来说，对于正则表达式中的每个操作（如连接、选择和闭包），都有对应的NFA构造方式，反之亦然。

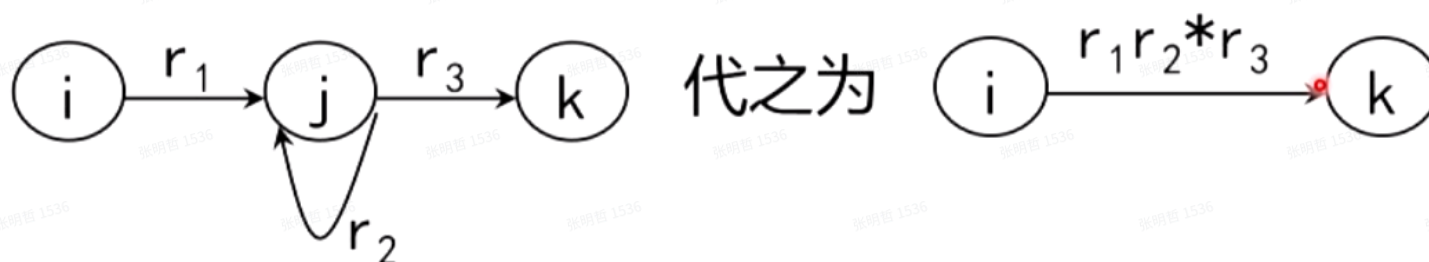
这个定理的存在为我们提供了一种将正则表达式转换为等价的NFA，并且可以利用NFA进行字符串匹配的方法，同时也为从NFA到正则表达式的转换提供了一种理论基础。

算法

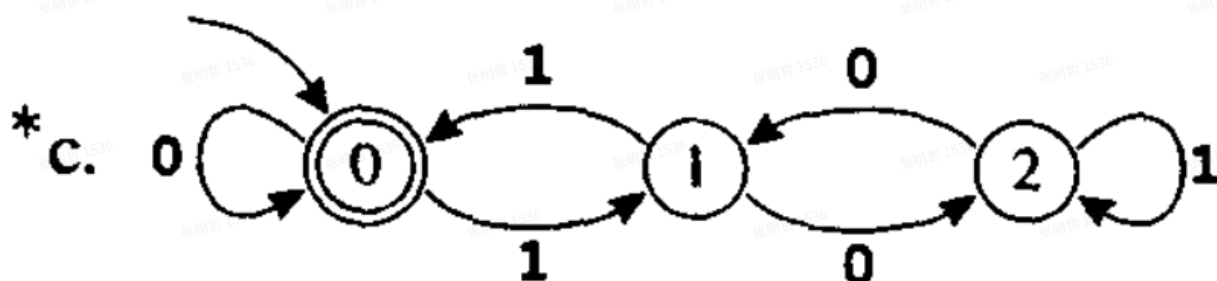
► 假定NFA $M = \langle S, \Sigma, \delta, S_0, F \rangle$ ，我们对M的状态转换图进行以下改造：

- 在M的转换图上加进两个状态X和Y，从X用 ϵ 弧连接到M的所有初态结点，从M的所有终态结点用 ϵ 弧连接到Y，从而形成一个新的NFA，记为 M' ，它只有一个初态X和一个终态Y，显然 $L(M) = L(M')$ 。





练习题

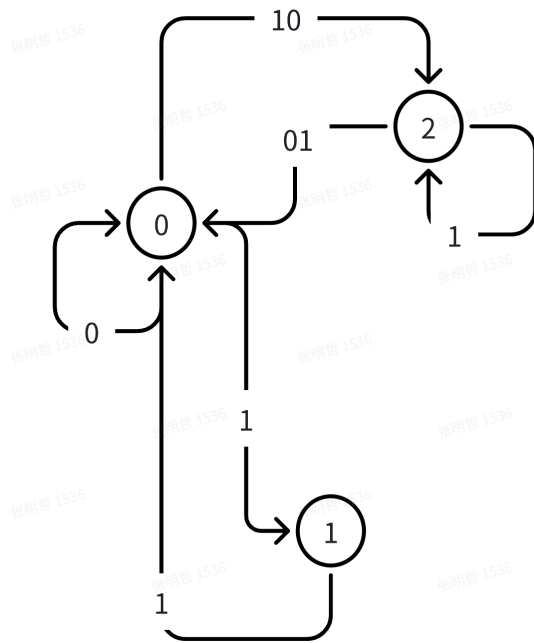


- *2.7 任何接收至少一个字符串的 DFA 都能转换为一个正则表达式。将习题 2.3c 的 DFA 转换为正则表达式。提示：首先假装状态 1 是初态。然后，编写一个通到状态 2 并返回到状态 1 的正则表达式和一个类似的通到状态 0 并返回到状态 1 的正则表达式。或者查看 Hopcroft 和 Ullman[1979] 一书中定理 2.4 关于此算法的论述。

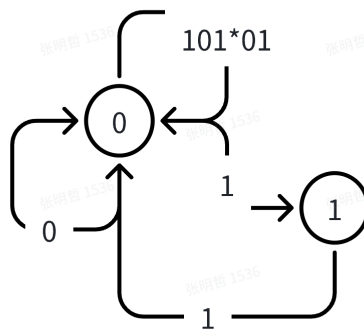
37

答案

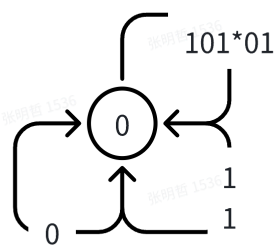
合并状态 0-1-2 和 2-1-0 后



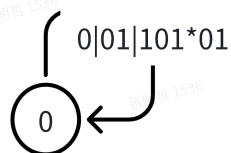
合并状态2左右和自身的克林子集后



合并状态0-1-0后



合并最后三个状态后



- 1 (0
- 2 |01
- 3 |101*01)*