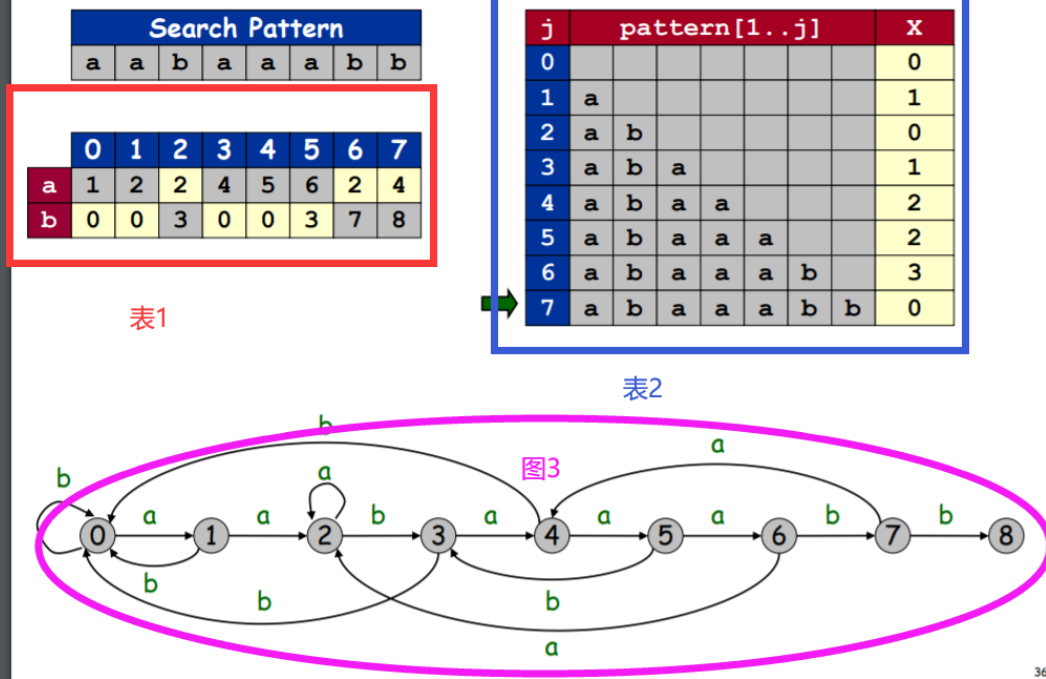


# FSA construction



问题 1:

表 1、表 2、图 3 是啥？

表 1 跟图 3 都是一个字符串和目标串匹配时的转移状态

表 2 的 x 列是原串的第 2 位至最后一位即  $P[2...n]$  和  $P[1...n-1]$  的按照转移状态表进行转移能到达的状态值

问题 2:

表 1 中，如果在第 i 个状态和原串匹配成功，则移动至状态 i+1。即

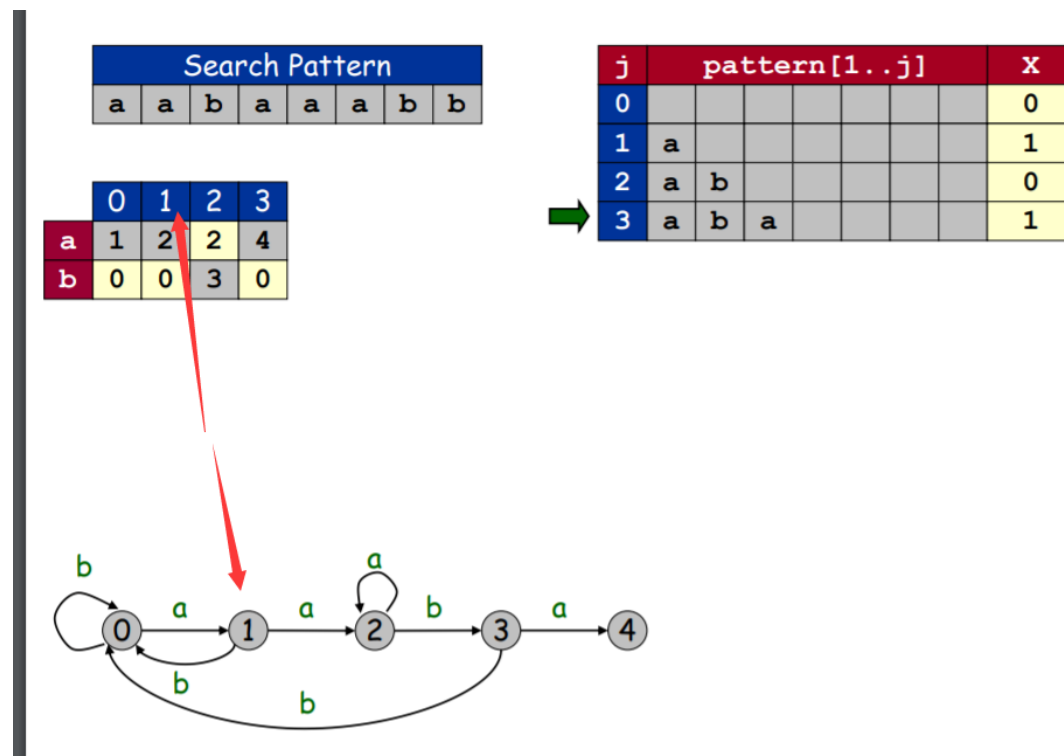
	0	1	2	3	4	5
a	1	2	?	4	5	6
b	?	?	3	?	?	?

那么如果第 i 个状态匹配失败应该移动到哪儿？

图中字符串为  $O="aabaaa"$ ，如果与  $S="aabab"$  进行匹配，进行到 S 的最后一位的 b 匹配失败了，我们希望把 b 与原串的第几位进行下一次匹配呢？

由于已经到第 5 位 b，因此我们已经可以得到  $O[0...4]=S[0...4]$  即  $aaba==aaba$ ，因此 我们考虑，将 b 移动到  $O[0...4]$  即 aaba 的最长公共前后缀的后面一位进行继续匹配。而  $O[0...4]$  的最长公共前后缀为 "a"，因此后面一位是第二个 "a"

如下图所示，匹配第二个 "a" 之前所在的状态为 1，因此我们的第五位 b 将从状态 1 开始进行后续的匹配，由于我们已经进行到状态 5 的匹配，下图中的都已经属于已知信息，因此从表 1 可得到状态 1 遇到 b 会进入状态 0。因此 aabab 应该进入状态 0。



### 问题 3:

如果失配要从最长公共前后缀所到达的状态进行下一步匹配,我要怎么知道公共前后缀会到达那个状态? ?

表 2 中  $P = \text{"abaaabb"}$  是字符串  $O = \text{"aabaaabb"}$  的最长真后缀  $O[2 \dots N]$ , 将  $P$  丢进图 3 进行状态转移可以到达一个某个状态  $k$  而到达状态  $k$  说明字符串  $P$  的第 (倒数第  $k$ ) 位到最后一位和字符串  $O$  的第 1 到第  $k$  位字符完全匹配, 因此因此  $O$  的第 1 到第  $k$  位即为  $O$  串的最长公共前后缀, 而状态  $k$  则是  $O$  的最长公共前后缀匹配成功后可到达的状态, 即我们问题 2 中所需要的状态。

观察表 2 可发现, 即为我们失配时所达到到的每一个状态

# FSA construction

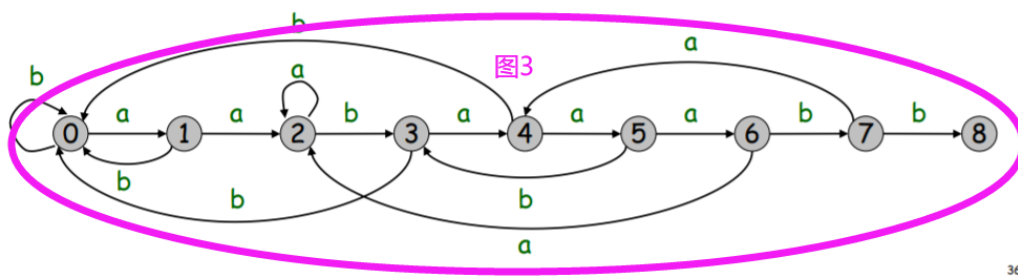
Search Pattern							
a	a	b	a	a	a	b	b

	0	1	2	3	4	5	6	7
a	1	2	2	4	5	6	2	4
b	0	0	3	0	0	3	7	8

表1

j	pattern[1..j]							x
0								0
1	a							1
2	a	b						0
3	a	b	a					1
4	a	b	a	a				2
5	a	b	a	a	a			2
6	a	b	a	a	a	b		3
7	a	b	a	a	a	b	b	0

表2



问题 4:

那这三个表，我们应该用什么顺序进行理解？

如下图所示

1. 假设我们已知状态 0-状态 4 的状态转移表 1 和图 3，我们需要获得状态 4 匹配 a 和匹配 b 的下一个状态，原串第 5 位为 a，因此状态 4 匹配 a 会进入状态 5
2. 那么，状态 4 匹配 b 应该进入哪个状态
3. 得到目前已经匹配的串 aaba，观察表 2，他的最长公共前后缀匹配到的下一个状态即他的最长真后缀 aba 匹配到的状态，即 aba 对应的行的 x 即状态 1
4. 观察表 1，得到状态 1 匹配 b 到达的状态为 0，因此表 2 的下一行 x 应为 0，而状态 4 匹配 b 的状态也是 0
5. 由此我们 O(1) 的得到了状态 4 匹配 a 和 b 所需要指向的两个状态。可以开始进入下一轮判断

# FSA construction

Search Pattern							
a	a	b	a	a	a	b	b

	0	1	2	3
a	1	2	2	4
b	0	0	3	0



j	pattern[1..j]							x
0								0
1	a							1
2	a	b						0
3	a	b	a					1

